



UNIVERSIDAD TECNOLÓGICA NACIONAL

INSTITUTO NACIONAL SUPERIOR DEL PROFESORADO TÉCNICO

En convenio académico con la Facultad Regional Villa María

LICENCIATURA EN LENGUA INGLESA

Tesis de Licenciatura

**THE INFLUENCE OF WRITTEN ITEMS IN THE
CONSTRUCTION OF GRAMMAR TESTS**

Tesista

PROFESORA JIMENA BUEDO

Director de la Tesis

MAGISTER MARIANO QUINTERNO

2011



UNIVERSIDAD TECNOLÓGICA NACIONAL

LICENCIATURA EN LENGUA INGLESA

Dissertation

**THE INFLUENCE OF WRITTEN ITEMS IN THE
CONSTRUCTION OF GRAMMAR TESTS**

Candidate

JIMENA BUEDO

Tutor

MAGISTER MARIANO QUINTERNO

2011

Dedication

To Mariano and Felicitas, whose support and readiness to help I will always be grateful for.

To my best teacher- my mother.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my tutor Magister Mariano Quintero for his insightful comments, his dedicated attention to detail and continuous support and encouragement.

A special mention to Dr. Omar Villarreal for his close follow-up, his useful advice and his respect for my timing in this process.

I would like to thank the students of Scania S.A, Carrier S.A, Motorcam S.A, Newport Cargo S.A and their teachers of English for being part of this study.

They have all rendered to this investigation any value it may have; the weaknesses in it are all my own responsibility.

Jimena Buedo

July 2011

Abstract

Testing, as McNamara (2000) expresses, is a universal feature of social life and as such it attempts to examine how people perform in relation to others to make decisions about their social roles. This concept also applies to language testing, which involves gathering information and making judgement within a context of specific purposes and goals (Frendo, 2009). Testing is of critical importance in shaping the teaching process and the decisions about the students' performance. Therefore, this research aimed to find out to what extent the test results obtained by in-company intermediate students were influenced by the type of written item chosen by the teacher in the construction of grammar tests. The study also explored whether previous training in class affected the test takers' performance. The difference between closed-items and open tasks, between items presented in isolation and in context was also highlighted in the hope of raising awareness of the range of results that might be attained with different test formats. Findings showed that the type of test item chosen to evaluate in-company intermediate students' knowledge of grammar in the written medium appeared to have influenced their test results.

Key words: assessment, grammar, in-company and test item.

Resumen

La evaluación, como lo expresa McNamara (2000), es una característica universal de la vida social y como tal intenta examinar como las personas se desempeñan en relación a otros para tomar decisiones acerca de sus roles sociales. Este concepto de evaluación también se aplica a la evaluación de la lengua la cual busca reunir información y hacer una valoración dentro de un contexto dado ya que los resultados logrados son importantes no solo por su influencia en la enseñanza sino también por las decisiones que se toman en base a ellos. Esta investigación intentó descubrir hasta qué punto el tipo de formato utilizado por los docentes para evaluar los conocimientos de gramática de alumnos intermedios estudiando inglés en empresas influye en los resultados obtenidos. Este estudio también exploró si la práctica previa de diferentes ejercicios en clase afecta el desempeño de los alumnos en el examen. Se enfatizó la diferencia entre distintos tipos de ejercicios con el objetivo de reflexionar sobre la variedad de resultados que pueden ser obtenidos al evaluar con diferentes formatos. Este estudio demostró que los ítems elegidos para evaluar el conocimiento de gramática de alumnos intermedios de inglés en

forma escrita parecen haber influenciado los resultados obtenidos.

Palabras claves: evaluación, gramática, cursos en empresas, ítems de evaluación

TABLE OF CONTENTS

TABLE OF CONTENTS	18
CHAPTER 1	19
INTRODUCTION.....	19
1.1. <i>Organization of this work</i>	22
1. TEST AND TESTING.....	24
1. a. <i>What is a test?</i>	24
1. b. <i>Testing the test</i>	25
2. TYPES OF TESTS.....	27
2. a. <i>Closed-item tests</i>	30
2. b. <i>Open-item tests</i>	32
2. c. <i>The construction of tests</i>	36
3- GRAMMAR TESTS.....	40
3. a. <i>Conceptions of grammar and their influence in testing</i>	41
3. b. <i>Testing grammar</i>	45
3. c. <i>Grammar test items</i>	46
4. TESTING IN-COMPANY GROUPS.....	51
4. a. <i>In-company students</i>	51
4. b. <i>In-company assessment</i>	53
CHAPTER 3	56
THE STUDY.....	56
3. 1. <i>The context of the study</i>	56
3. 2. <i>Procedure and research methods</i>	59
METHODS.....	61
3. 3. a. <i>Grammar test</i>	61
3. 4. <i>Questionnaire</i>	75
3. 5. <i>Semi-structured interview</i>	84
CHAPTER 4	90
THE RESULTS OF THE STUDY.....	90
1- <i>Grammar test</i>	90
2- <i>Questionnaire</i>	109
3- <i>Teachers' interview</i>	117
TRIANGULATION OF RESEARCH RESULTS.....	125
CHAPTER 5	138
CONCLUSIONS.....	138
5.1. <i>Limitations and suggestions for future research</i>	142
REFERENCES.....	144
APPENDIX I.....	147
APPENDIX II.....	151
APPENDIX III.....	153
APPENDIX V.....	156
APPENDIX VI.....	156

"It is a fine thing to have ability but the ability to discover ability in others is the true test."

(Lou Holtz)

CHAPTER 1

INTRODUCTION

Throughout history, people have been put to the test to show their performance in a particular field and to control entry to many important social roles not only in education but also in the workplace (McNamara, 2007). What is true of testing in social life is also true of language testing since the latter plays a powerful role in education, employment and traveling (McNamara, 2007).

There seems to be a deep mistrust towards tests and testers simply because language tests often fail to measure accurately what they intend to measure (Hughes, 2003) and although teaching is the primary activity, testing is necessary because it shapes the teaching process (Frendo, 2000).

Bachman (2004) explains that language tests have become a pervasive part of education because the scores from language tests are used to make inferences about a student's language ability and consequently decisions are made based on those results.

When a test is used, it is important to define whether it will measure an aspect of language ability, some kind of progress in language learning or the use of language in real-life settings (Bachman, 2004).

Evaluating a one-off, in-company course, for example, will differ from evaluating a university course that runs many times a year (Frendo, 2009). Business English courses may be evaluated at different levels and from different perspectives since the evaluation can focus on teaching techniques, on the reaction of the participants towards the content of the course, on reflective approaches to improve the learning process or on the customer's requirements to analyze the benefits of the course (Frendo, 2009).

As there are so many factors that contribute to the construction and effectiveness of a test and important decisions are made based on the learners' test results, the following question is brought to light: Are teachers aware of the importance of designing or choosing appropriate test items to measure their students' performance?

This investigation, therefore, is informed by the following research question:

To what extent does the type of test item chosen to evaluate grammar in the written medium influence in-company intermediate students' test results?

As derived from the research question, the following hypotheses will also be examined:

1-Close-item exercises seem easier than open-item tasks.

2-Close-item exercises in a given context may appear clearer for the students to complete than exercises in isolation.

3-Close-item tasks seem to require previous practice rather than knowledge.

4-Open-item tasks aimed at checking knowledge of grammar may not always reflect the use of intended grammar structures.

1.1. Organization of this work

This dissertation is made up of five chapters. Chapter 1 includes the introduction, the research question and the hypotheses derived from it, and the description of the organization of this work.

Chapter 2 deals with the literature review. The concept of *testing* is defined and the importance of testing a test is brought into consideration. Afterwards, the different type of test items -closed and open- together with the considerations for item test construction are focused on. The chapter ends with a definition of grammar and different tasks that may be used to check students' grammar performance along with a thorough description of intermediate in-company students and the different ways in which they are assessed.

Chapter 3 engages in the present investigation: the analysis of the context of the study and the description of the procedure and research methods.

Chapter 4 portrays the results of the study. The data provided by each of the instruments of the investigation are analyzed in isolation in the first three sections of chapter 4. The last section in this chapter deals with the triangulation of the results.

Chapter 5 displays the conclusions of the study, its limitations, the suggestions for future research and a reference section with the list of works cited.

For the sake of this investigation, the words *error* and *mistake*, *assessment*, *test* and *evaluation*, *objective* and *closed items* were used as synonyms regardless of the difference in meaning that some authors might have established.

CHAPTER 2

LITERATURE REVIEW

1. Test and testing

1. a. What is a test?

For years, tests have been part of any learning environment with the aim of measuring students' knowledge in a particular field (Bachman, 2004). Many definitions can be found to explain what a test is and the purpose it has.

The *Cambridge International Dictionary of English* states that a test is “a way of discovering, by questions or practical activities, what someone knows, or what someone or something can do or is like” (1996:1505). In the light of this definition, tests aim at discovering something. Similarly, language testing attempts to find out about a person's language abilities. In McNamara's words, a test is “a procedure for gathering evidence of general or specific language abilities from performance on tasks designed to provide a basis for predictions about an individual's use of those abilities in real world contexts” (2007:11).

In accordance with the definitions below, Frenco (2009) considers that tests involve asking questions, collecting relevant information, and making judgment within a context of specific purposes and goals. A careful look at the similar ways the concept of testing is defined shows how emphasis is repeatedly

placed on the need to discover people's abilities by asking questions or designing activities within a certain context.

1. b. Testing the test

Test-designers should not only construct the test bearing in mind the use for which it is intended, but they should also provide a measure that can be interpreted as an indicator of an individual's language ability. Bachman and Palmer (1996) consider elements such as reliability, validity, authenticity, interactiveness, impact and practicality essential to the usefulness of any language test.

Reliability is defined as a consistency of measurement. Human beings simply do not seem to behave in exactly the same way on every occasion, even when the circumstances seem identical. One should construct, give and grade tests in such a way that the scores obtained on a particular occasion are very similar to those which would have been obtained by the same group of students, with the same level at a different time (Hughes, 2003). To minimize the effects of inconsistencies, Hughes (2003) explains that the more items, the more reliable the test.

Validity, on the other hand, refers to the appropriateness of a test as a measure of what it is supposed to measure. A test is valid; therefore, if it reflects the area of language ability we want to measure (Bachman and Palmer, 1996).

Alderson *et al* state that internal validity can be assessed in terms of face validation, content validation and response validation. Face validity involves an intuitive judgment about the content of the test by non-expert people such as students or administrators (Alderson Clapham and Wall, 2005). Content validity involves gathering the judgement of experts who are to be trusted and response validity refers to the way individuals respond to test items and the reasoning they go through in order to respond (Alderson Clapham and Wall, 2005).

External validity refers to two types of validation: concurrent validity, which involves the comparison of a candidate's test score with some external measure taken at the same time or predictive validity, which compares scores some time after the test was given (Alderson Clapham and Wall, 2005).

A test should also be authentic so, in Bachman and Palmer's (1996) words, there should be a correspondence of the characteristics of a given language test task and the characteristics of a target language use task.

As to interactiveness, as Bachman and Palmer (1996, p.39) express: “It pertains to the degree to which the constructs we want to assess are critically involved in accomplishing the test task.” It reflects the type of involvement of the testees’ language ability to accomplish a test task (Bachman and Palmer, 1996).

When it comes to testing the students’ grammatical knowledge, for a test to be useful, it needs to be constructed with a specific purpose in mind. Purpura (2004) explains that once the purpose of a test has been clearly stated and the language-use tasks have been identified, the major challenge to consider is to define what areas of grammatical knowledge the examinee needs to display so that the test measures what it intended to. In most testing contexts, the definition of test construct derives from a textbook, syllabus or set of objectives. However, before the test is put to operational use, teachers should decide how much evidence needs to be gathered to demonstrate that the test-taker has grammatical ability (Purpura, 2004).

2. Types of Tests

Not all tests are of the same kind since they differ with respect to their design and purpose. With respect to their purpose, tests can be classified into *achievement tests*, which establish how successfully students have achieved objectives as in the case of *final achievement tests*, given at the end of a

course of study and *progress achievement tests*, which measure students' improvement (Hughes, 2003). Then there are *entry or placement tests*, which indicate the level of a student while *diagnostic tests* are used to find out problem areas. Last but not least, *proficiency tests* attempt to give evidence of the students' ability in a foreign language (Hughes, 2003).

Tests not only differ on their purpose but also on their method. McNamara (2007) distinguishes two types of test methods: *the traditional paper-and-pencil tests* and *performance tests*. In *performance base tests*, language skills are assessed in an act of communication whereas test items in traditional tests are generally presented in a fixed response format from which the candidate is required to choose the right alternative out of a given number of options (McNamara, 2007).

There seems to be a tendency to test aspects of knowledge of the grammatical system in isolation with minimal context. McNamara (2007) refers to discrete-point (or objective) testing when the response mode is a matching, true-false, multiple choice, or fill-in-the-blank format, in which a response is either selected from among different alternatives provided or otherwise restricted by the nature of the context provided.

Although in the 1960s the psychometric-structuralist period was considered the most suitable method, this discrete point tradition of testing was also seen as

focusing exclusively on knowledge of the formal linguistic system rather than on the way such knowledge was used to achieve communication (McNamara, 2007).

For Heaton (1988), objective items such as multiple-choice tasks or completion items are used to test the ability to recognize or produce correct forms of the grammatical features of the language rather than the ability to use language to express ideas. However, it has been proved that when properly designed, tests of this type require the student to use appropriate grammatical structures rather than simply recognize their correct use (Hughes, 2003).

If the aim is to assess grammar, Purpura (2004) categorizes tasks for grammar tests according to the type of response. There are three types of tasks: *selected-response*, *limited production* and *extended production tasks*.

In *selected-response tasks*, test-takers are expected to select a response presented in the form of an item. These are multiple-choice, error identification, matching, discrimination and noticing tasks. (Purpura, 2004) *Limited-production tasks* elicit a short response which is intended to assess one or more areas of grammatical knowledge. The range of possible answers can be larger than in selected-response tasks. The gap-filling task, the short-answer task and the completion task are some examples of limited production (Purpura, 2004).

2. a. Closed-item tests

Even though there are several types of fixed response format, for McNamara (2007), the most important kind is *multiple choice*. Multiple choice items, as Hughes (2003) explains, take many forms but their basic structure includes a stem, which can be a sentence, phrase or passage and a number of options, in most cases, one of which is correct or most appropriate, the others functioning as distractors.

Alderson, Clapham and Wall (2005) distinguish other objective-type items such as dichotomous items, matching items, ordering tasks and editing items. Dichotomous items are true or false items, matching items are often used in reading and listening comprehension tests, where learners have to transfer information to a chart, table or map. (Alderson, Clapham and Wall, 2005). In ordering tasks, candidates are asked to put a group of words, sentences or phrases in order and in editing items students have to identify the errors which have been introduced in a set of given sentences or passages. (Alderson, 1992)

With respect to the types of item choice tests, Oller (1989) proposed cloze tests, where the learner is expected to integrate grammatical, lexical, contextual and pragmatic knowledge in a more economical way since words or phrases are deleted from a short passage and test takers have to restore the missing

words, which are usually aspects of language such as grammar or vocabulary (Alderson, Clapham and Wall, 2005).

Alderson (2005) also identifies the cloze test which, as mentioned above, refers to tests in which words are deleted mechanically -every sixth word is removed from a passage regardless of its function. There is also the C-test format, which also involves the mechanical deletion of words from a text, but in this case, the first few letters of the missing word are provided, thus reducing the number of possible answers.

These closed item tasks seem to be quite artificial and they might not test students' correct use of grammatical structures (Hughes, 2003). Still, this depends on how they are designed and what their purpose is. Heaton (1988) explains that *completion items* are a useful means of testing students' ability to produce acceptable and appropriate forms of language on the grounds that they measure production rather than recognition. They test the ability to insert the most appropriate grammatical or functional words in selected blanks in the sentence instead of using items in isolation. A passage of continuous prose in these cases might prove to be useful because, according to Heaton (1988), it would avoid ambiguity on the one hand and, on the other, the student would be required to pay attention to all the context clues in the process of elicitation.

The transformation type of item, which tests the ability to produce structures in the target language, may measure some of the writing skills (Heaton, 1988). This is a good device for testing students' knowledge of tenses and verb forms; by changing words or constructing broken sentences, students may show their ability to write full sentences.

2. b. Open-item tests

“An open-ended item is one which does not require a specific answer (Frendo, 2000:131)”. This format requires the candidate to combine many language elements for the completion of the task such as writing a composition or making notes while listening to a lecture (Hughes, 2007). These items have the advantage of not constraining the test taker as in closed-items but decisions need to be made about the acceptability of responses (McNamara, 2007).

The need of assessing the practical language skills of foreign language learners wishing to study abroad led to a demand for language tests which involved integrated performance on the part of the test taker (McMamara, 2007). Purpura (2004) refers to these items as *extended-production tasks* since they present input in the form of a prompt which aims at eliciting large amounts of data such as information-gap, role-play and simulation tasks.

From the early 1970s, Hymes's theory of communicative competence began to exert influence on language teaching and language testing in the hope of engaging candidates in an act of communication and assuming roles in real world settings (McNamara, 2007).

The open item format has the advantage of reducing the effect of guessing so the candidate should assume greater responsibility for the response and this may be more demanding and more authentic but, at the same time, more difficult to score since agreement on what constitutes an acceptable response needs to be achieved (McNamara, 2007).

At first sight, writing the prompts for written composition seems very easy since all one seems to have to do is write a topic for the student to answer (Alderson Clapham and Wall, 2005). However, as Alderson *et al* (2005) state, candidates need to know how long the essay should be, for whom the essay is to be written and if they will be marked on accuracy, the ability to produce a good argument or solely for the use of grammar and vocabulary. Therefore, the design of written tasks should involve gathering information about the test purpose, characteristics of the target population and their real-world needs (Weigle, 2009).

As Butler *et al* (1996) explain, "The test taker will generate a written response of a paragraph or more in length to a prompt of two or more sentences that

explicitly specifies the elements of the writing task including audience, purpose and source of information content” (as cited in Weigle, 2009:78).

If students are asked to write an essay or a composition, the question should include not only the expected length, the mark that will be given for the structure of the essay, the clarity of argument, the range of grammar and vocabulary used but also a short piece of text which sets the scene or provides background information if creativity is expected (Alderson Clapham and Wall, 2005).

Summaries are used mainly to test reading or listening comprehension but they may closely replicate many real-life activities (Alderson, Clapham and Wall, 2005). According to Alderson et al (2005) many tasks are not as formal as essays and compositions since test takers are sometimes asked to write an informal letter or note, in which case, it may be necessary for the item writer to invent a scenario which would require the candidate to write in the foreign language.

A written prompt may include source materials such as a reading passage, a brief quotation, a drawing or it can simply present a topic (Weigle, 2009). Providing stimulus material allows students to focus on the linguistic aspects of the task and it also provides a common basis of information for all test takers to activate their background knowledge or schemata (Weigle, 2009).

The ability of writing effectively is assuming an important role globally and instruction in writing is essential in second and foreign language education for if compared to speaking, writing seems a more standardized system which must be acquired through special instruction (Weigle, 2002). Thus, as the role of writing in education is increasing, there is greater demand for ways to test writing successfully (Weigle, 2002). In Hughes's words, "the best way to test peoples' writing ability is to get them to write" (1989:75). Weigle (2002) suggests the following key points before designing assessment tasks: what we are trying to test, whether we are interested in grammatical sentences or a communicative function, what criteria or standards will be used and what constraints can limit the information collected about test takers' writing ability.

As Weigle (2002) explains, the traditional role of writing in a language class is to reinforce the knowledge about the structure and vocabulary of the language but looking beyond the classroom to the real world, learners may need to write for informational purposes from a report form to a letter of application for a job. Students might need to reproduce information which has already been determined, to organize information known to the writer or to create new information (Weigle, 2002). Vahapassi (1982) also lists six dominant purposes for writing which are: to learn, convey emotions, inform, convince, persuade and entertain (Weigle 2002).

Writing is, as Hayes (1996) states, not only the product of an individual, but also a social and cultural act since what we write, the way we write and who we write to is shaped by social conventions. Weigle (2002: 22) agrees that “the ability to write indicates the ability to function as a literate member of a particular segment of society or discourse community.” Therefore, learning to write seems to involve much more than simply learning the grammar and vocabulary of another language. Describing the various influences on the writing process is thus an important step when developing or using a test of writing not only to help define the skills tested more clearly but also to find out about other influences that may affect writing but which are not related to what is being assessed (Weigle, 2002).

2. c. The construction of tests

Oller (1989) lists a number of steps that test designers should take into account, such as obtaining a clear notion of what it is that needs to be tested, selecting appropriate item content and devising a suitable item format, writing the test items, getting some qualified person to read the test items, rewriting any weak item, pretesting the items, running an item analysis over the data from the pretesting and discarding non-functional alternatives.

According to Heaton (1988), all multiple-choice items should be appropriate to the level of the candidates, so the context in which the item is introduced

should be at a lower level than the item which is being tested. Each item should be as brief and clear as possible and it is important to have some simple items at the beginning, especially if candidates are not very familiar with the test format.

The most important requirements for these item tests, in Alderson's (1992) view, are that the correct answer must be "genuinely" correct to avoid problems such as giving two acceptable alternatives. In addition, the presentation of the context should be very clear since it might be the case that the item writer might have one context in mind which may not be so obvious to the test taker. The stem of each item should, therefore, convey enough information to indicate the basis from which the correct option should be selected and each distractor should appear right to any student who is unsure of the correct option. The purpose of this type of tasks, in Heaton's (1988:60) words, is that "testees should select their option rather than eliminate the incorrect options."

In tests with a number of objectively scored test items, it is usual to carry out a procedure known as item analysis, which tells us if test items are at the right level for the group or if the items provide useful information consistent with the one provided by the other items (McNamara, 2007).

Some authors, such as Hughes (2003), identify some drawbacks in closed item selection tests. A multiple choice grammar test score, for example, may be a

poor indicator of someone's ability to use grammatical structures. The person who can identify the correct response in the item above may not be able to produce the correct form when speaking or writing. This is, in part, a question of construct validity; whether or not grammatical knowledge of the kind that can be demonstrated in a multiple choice test resembles the productive use of grammar. Even if it does, there is still a gap to be bridged between knowledge and use; if use is what we are interested in, that gap will mean that test scores are at best giving incomplete information.

However, as Heaton (1988) explains, there are occasions when good objective grammar tests may be useful as long as such tests are not regarded as a measure of the students' ability to communicate in the target language since closed-items do not lend themselves to the testing of language as communication. So, even though closed item exercises rarely measure communication as such, they can prove useful in measuring students' ability to recognize correct grammatical forms and to make important discriminations in the target language and identify areas of difficulty (Heaton, 1988). The number of alternatives for each multiple-choice item is five in most tests but a larger number might reduce even further the element of chance, the test must then be long enough to allow for a reliable assessment and short enough to be practicable (Heaton, 1988).

Another drawback which closed- item test designers should try to overcome is that there seems to be evidence that students taking multiple-choice tests can learn strategies for guessing the correct answer, for eliminating distractors, for avoiding two options that are similar in meaning or for selecting an option that is longer than the other distractors (Alderson, 1992). However, as Heaton (1988) explains, four or five alternatives for each item are sufficient to reduce the possibility of guessing. Furthermore, experience shows that candidates rarely make wild guesses since most testees base their guesses on partial knowledge.

Another criticism that Hughes (2003) points out is that multiple choice items require distractors and in a grammar test, for example, it may be difficult to find three or four plausible alternatives to the structure we intend to test. Therefore, there may be clues in the option as to which the correct answer is and the distractors may end up being ineffective (Hughes, 2003). Therefore, it seems that as Bachman and Palmer clearly (1996:40) express: “Good multiple choice items are notoriously difficult to write and always require extensive pretesting.”

Thus, objective tests require far more careful preparation than subjective tests because even if these are frequently criticized on the grounds that they are simpler to answer, items in an objective test can be made just as easy or as difficult as the test constructor wishes (Heaton, 1988).

According to Bachman (2004), tests and test scores are used to make inferences about individuals' language ability, to make decisions about students and to help teachers collect useful information. Still, we need to be able to demonstrate that the scores we obtain from language tests are reliable, and that the ways in which we interpret and use language test scores are valid (Bachman, 2004). But the scoring will also depend upon the teachers' approach to testing. Bachman (2004) considers two approaches: the counting approach, which is used most typically with tasks in which individuals select a response from among several choices that are given, respond along points on a scale, or with items that require completion or short-answers; and the judging approach, which is frequently used with tasks that require test takers to produce an extended sample of language, such as in a composition or an oral interview.

Test scores can also be influenced by many factors such as the test-takers' age, gender, background, motivation, level of anxiety or because of the test itself (Bachman, 2004). In fact, as Purpura (2004:100) explains, "the type of questions on the test can severely impact performance." Thus, it is important for test developers to understand the characteristics of each task they use to determine the test-taker's grammatical ability since some, for example, seem to perform better on multiple-choice tasks than on oral interviews and others do better on essay writing than on cloze tasks. (Purpura, 2004)

3- Grammar tests

3. a. Conceptions of grammar and their influence in testing

There have been different definitions and conceptualizations of grammar over the years. When most second language researchers, teachers and testers think of “grammar”, they think of the study of the language derived from data taken from native speakers (Purpura, 2004).

Most linguists have adopted either a syntactocentric perspective of language, where syntax is the central feature to be analyzed; or a communication perspective, where the emphasis is on how language is used to convey meaning (Purpura, 2004).

According to the syntactocentric view, formal grammar is defined as “a systematic way of accounting for and predicting an ‘ideal’ speaker’s or hearer’s knowledge of the language, which consists of a set of rules or “principles” that can be used to generate all well-formed or grammatical utterances in the language” (Purpura, 2004:6). This approach is concerned mainly with the structure of sentences rather than the literal and contextual use. Purpura (2004) believes that most classroom language teachers base their syllabus design, instruction and assessment on grammatical forms and the rules that govern them.

The second general approach views language as a system of communication where grammatical forms are used to convey a number of meanings (Purpura 2004). This perspective focuses on the overall message and the interpretations that it might invoke. These two general approaches have influenced assessment since they are central to determining how grammar is conceptualized in language teaching and learning (Purpura, 2004).

Unlike traditional grammars, structural grammars describe the structure of a language in terms of its morphology and syntax. Another well-known syntactocentric theory is Chomsky's (1965) universal grammar, which seeks to provide a "universal description of language behaviour through a set of phrase-structure rules, which describe the underlying structure of all languages" (Radford 1997:265). However, Chomsky's theory has been criticized for excluding the role of pragmatics or meanings derived from context use (Purpura, 2004).

In the grammar-learning process, explicit grammatical knowledge refers to a conscious knowledge of grammatical forms through the explanation of a rule whereas implicit grammatical knowledge involves semantic processing of the input without a rule presentation (Purpura, 2004). According to Purpura (2004), the majority of studies surveyed showed that explicit grammar instruction helps language learners achieve higher levels of grammatical ability.

In 1980, Canale and Swain defined grammatical competence as knowledge of the rules of phonology, the lexicon, syntax and semantics but no explanation was provided on how their framework accounted for cases in which grammar was used to express meanings beyond sentence level.

In the past, as Rutherford (1988) defines, grammar was the analyses of a language system and this was thought to be sufficient for learners to actually acquire another language. Purpura (2004) explains that language teachers today would maintain that grammar is an indispensable resource for effective communication but not an object of study in itself.

Rea-Dickins (1991) stated that grammar is the embodiment of syntax, semantics and pragmatics and that the goal of communicative grammar tests is to provide an opportunity for the test-taker to create a message and to produce appropriate grammatical responses in a given context. (Purpura, 2004)

Bachman and Palmer (1996) viewed language ability as an internal construct, consisting of language knowledge that interacts with internal characteristics as well as with the features of the language-use context.

For Bachman and Palmer (1996) there are two general components to describe language knowledge: organizational knowledge, which is how individuals

control language structure to produce grammatically correct utterances, sentences and texts. This component is divided into grammatical and textual knowledge. The former is the learner's knowledge of vocabulary, syntax and phonology (how sentences are organized) and the latter refers to knowledge of cohesion, rhetorical and conversational organization (how sentences are organized into texts) (Purpura, 2004).

The second component is pragmatic knowledge, which is defined as an individual's ability to communicate meaning and produce contextually appropriate utterances, sentences and texts. This second component refers to functional knowledge, which is how an individual expresses or interprets language in communicative settings; and pragmatic knowledge, which, in sociolinguistic terms, refers to how language is used contextually. (Purpura, 2004)

According to Larsen-Freeman (1997) grammatical knowledge can be categorized along three dimensions: linguistic form, semantic meaning and pragmatic use. These three dimensions may be independent or interconnected but the boundaries among them are not always distinct (Purpura, 2004).

Purpura (2004) states that grammatical knowledge embodies two highly related components: grammatical form and grammatical meaning. Knowledge of grammatical form includes phonological, lexical, morphosyntactic, cohesive,

information management and interactional levels whereas grammatical meaning refers to the literal meaning expressed by words, phrases and sentences. The notion of meaning is a critical component in the assessment of grammatical knowledge since the main goal is usually to determine if learners are able to use forms to express themselves accurately (Purpura, 2004). Therefore, it is important for testers to make a distinction among the components mentioned above so that assessment can be used to provide more precise information about students' knowledge of grammar (Purpura, 2004).

Purpura (2004:84) makes a distinction between the terms knowledge and ability: "Knowledge refers to internalized informational structures that are built up through experience and stored in long-term memory" and grammatical ability involves the capacity to use these informational structures meaningfully, combining grammatical knowledge and strategic competence (Purpura 2004).

3. b. Testing grammar

Grammar has been central to language assessment for many years. Language teachers have always acknowledged the link between teaching and testing and have always assessed their students' knowledge of grammar (Purpura, 2004). Still, it seems that there are different perspectives to testing students' grammatical knowledge. Purpura (2004) claims that many of the grammar tests

that are currently in use reflect the perspectives of structural linguistics and discrete-point measurement.

Throughout time, grammar has been assessed in different ways. Purpura (2004) distinguishes different ways of testing it. It can be done through the ability to recite rules, through the ability to extrapolate a rule from samples of the target language, or through the ability to provide an accurate translation. Knowledge of grammar might also be inferred from the ability to select a grammatically correct answer from several options on a multiple choice test, to supply a grammatically accurate word or phrase in a paragraph or dialogue, to construct grammatically appropriate sentences, or to provide judgements regarding the grammaticality of an utterance. It seems that all instances of language use invoke the same fundamental working knowledge of grammar and the lack of it severely limit what is understood or produced in communication.

3. c. Grammar test items

Given the many ways of interpreting what it means to know grammar, the design of a test should reflect the different components of grammatical knowledge, the purpose of the assessment and the learners about which inferences are made. Regardless of the assessment purpose, it is important to

be precise in the definition of grammatical knowledge for the construction and validation of tests (Purpura, 2004).

Loschky and Bley-Vroman (1993) identified three types of grammar tasks: task-naturalness, task utility and task essentialness. The first involves a task where a grammatical construction may be used naturally or it may be performed well without this grammatical construction. The second, task utility, refers to the completion of a grammar task without using the most appropriate structure and the third is task essentialness which implies that the item cannot be completed unless the grammatical form is used (Purpura, 2004).

Purpura (2004) explains the importance of designing tasks to elicit scorable grammatical performance:

“As the goal of grammar assessment is to provide as useful a measurement as possible of our students’ grammatical ability, we need to design test tasks in which the variability of our students’ scores is attributed to the differences in their grammatical ability, and not to uncontrolled or irrelevant variability resulting from the types of tasks or the quality of the tasks that we have put on our tests.” (Purpura, 2004:112)

The effects of teaching and learning seem to present a number of challenges for language testing. Ellis (2001:45) states that “the notion that grammatical

knowledge structures can be differentiated according to whether they are fully automatized (i.e., implicit) or not (i.e., explicit) raises important questions for the testing of grammatical ability.” There are many purposes for assessing our students’ knowledge of grammar and we might assess it explicitly or implicitly. If we might want to test the learners’ explicit knowledge of one or more grammatical forms, the learner could be asked to answer multiple-choice or short-answer questions (Purpura, 2004). The information taken from these tests would show if the students can apply certain grammatical forms regardless of whether they can use them in fluent and spontaneous discourse (Purpura, 2004). This type of assessment could be useful for teachers wishing to determine if their students have mastered certain grammatical form. If the objective is to assess the internalization of a grammatical form in spontaneous use, a task to elicit comprehension or full production should be carried out (Purpura, 2004).

For Carroll (1968), grammatical competence incorporates the morphosyntax and semantic components of grammar; therefore, tests should be designed to predict the use of language elements in future social situations. This is why discrete-point tasks should be, according to Carroll, complemented by integrative tasks that would also assess the degree to which the learner can use his/her knowledge of grammar in real-life tasks (Purpura, 2004).

Oller (1979) proposes, instead, a view of second or foreign language proficiency in terms of an individual's "pragmatic expectancy grammar," which is defined as a system that learners use to relate sequences of linguistic elements that conform to the contextual constraints in real life messages (Oller, 1979). A gap-filling task may illustrate the notion of "expectancy grammar." Here the test-takers read a passage and by interpreting the context, they may predict the information for the gaps by relating to linguistic form, semantic meaning, and/or pragmatic use, or their rhetorical or sociocultural knowledge (Purpura, 2004).

The following example is given:

"A test-taker might examine the linguistic environment of the gap and determine from the sequential organization of language (i.e., expectancy grammar) that a verb best completes the gap. He or she might also decide that the verb needs to carry past meaning and embody a specific lexical form. Finally, in realizing that the contextual focus of the sentence is on the action and not on the agent, the test-taker uses a passive voice construction (pragmatic use). In sum, pragmatic expectancy grammar forces the test-taker to integrate his or her knowledge of grammar, meaning and pragmatic use to complete the task" (Purpura, 2004:52).

Grammar, in Oller's (1979) view, is an integration of linguistic form and pragmatic use within a given context. If our aim is to measure linguistic forms alone- for example, the present perfect tense forms- a discrete-point test of grammar could be constructed. However, it would not account for the different types of meanings that grammatical forms encode. As Purpura (2004:56) clearly expresses: "If our assessment goal were limited to an understanding of how learners have mastered grammatical forms, then the current models of grammatical knowledge would suffice." However, if teachers hope to understand how students use grammatical forms to convey different meanings, the dimensions of grammatical ability should be considered (Purpura, 2004).

Bachman and Palmer (1996) organize test development into 3 stages: design, operationalization and administration.

The design of a test should contain the following components: the purpose of the test (progress, achievement, selection, placement, proficiency), the types of tasks, a description of the test-takers, the constructs to be measured (the use of strategic competence, cognitive, metacognitive, social or affective strategies), a plan for evaluating its usefulness and how to deal with material and time resources (Purpura, 2004).

The operational stage of grammar-test development has to do with the structure of the test and the test-task specifications to provide congruence

between what is supposed to be on the test and what the test actually includes. The writing of the test also takes place in this phase (Purpura, 2004).

The last stage involves the administration and the analysis of results, which can be carried out by piloting the test to collect data and thus improve or support the contents of the test (Purpura, 2004).

4. Testing in-company groups

4. a. In-company students

Over the last years, Business English has emerged in the world of enterprises as an urgent need amid those who are part and parcel of the global economy (Ellis & Johnson, 2000). Business English, therefore, is taught in public or private institutions, in language schools, in adult learning centers and a great deal of training also takes place in companies with trainers who are employees of the company itself or with teachers coming from outside (Ellis & Johnson, 2000). However, not all courses run by a company or a business college necessarily merit the title of “Business English” since the term is used to cover a variety of “Englishes” (Ellis & Johnson, 2000). Business English should be seen in the overall context of English for Specific Purposes (ESP) but it is often a mix of specific content and general content so the course content may also be drawn from a General English course book to improve the participants’ general command of English (Ellis and Johnson, 2000).

Within the wide variety of situations in which people learn and use second languages, Weigle (2002) distinguishes three groups of adult second language learners. The first group consists of immigrants to a foreign country who may or may not be literate. The second group consists of those who seek an advance university degree and the last group is made up of adults learning a second language for personal interest or career enhancement (Weigle, 2002).

English has become the international language of Business and therefore non native speakers aim at communicating effectively rather than sounding like native speakers of English (Frendo, 2009).

Many learners attend English courses where groups are usually formed on the basis of their language level rather than their job (Evans and John, 2009). The underlying focus of these courses is often grammatical and there are activities which are more open-ended to develop fluency (Evans and John, 2009).

In contrast, English for Specific Business Purposes courses aim at job-experienced students who focus on specific communicative events which stem from the learner's own business context (Evans and John, 2009).

Many working people cannot take much time for language learning so they build language learning through telephone classes, tutored self-study, tutored

distance-learning on the internet or by listening to tapes while traveling (Evans and John, 2009).

4. b. In-company assessment

Evaluating an in-company course will differ in many aspects from evaluating a university course (Frendo 2009). One model of evaluating business English training proposed by Frendo (2009) is based on Kirkpatrick's five levels of evaluation which include the learners' reaction to the teaching, what was learned, the transfer of what has been learned to the workplace and the results of this training (Frendo, 2009).

Another way of evaluating in-company students is through the use of categories (Frendo, 2009). Summative evaluation intends to find out whether or not the course objectives have been achieved, formative evaluation focuses on making improvements for the future, and illuminative evaluation aims at reflecting upon the teaching and learning process (Frendo, 2009).

The evaluation of progress might depend on the aim of the course which is usually defined in relation to the skills required in the job or course of study. Formal examinations as Ellis and Johnson (2000:13) explain: "include a written paper in which marks are awarded for grammatical accuracy as well as range of vocabulary and appropriacy."

Institutions and teachers working in company premises might use an examination offered by a British examining body or the trainer might choose to assess the students' level independently (Ellis & Johnson, 2000).

There are published tests and examinations leading to certification in Business English such as the London Chamber of Commerce and Industry Examinations Board, the University of Cambridge Examination Syndicate and the University of Oxford Delegacy of Examinations. Adult students studying in-company sometimes take General English examinations such as the Cambridge First Certificate or Proficiency, the American-based TOEFL and TOEIC (Ellis and Johnson, 2000) or the BEC (Business English Certificate), which require test takers either to achieve a threshold score in order to pass or the score can be later compared to a set of can-do statements (Frendo, 2009).

If the trainer working for an organization has a different system of testing or works independently, it is important to determine the course objectives and to get hold of a scale to assess the learners' level (Ellis & Johnson, 2000). It might be simpler to assess the learner's knowledge of grammar by setting a multiple choice, itemized test of the traditional type. Still, there is often a mismatch between the learner's knowledge of grammar and the level of communicative ability (Ellis and Johnson, 2000). Therefore, one way to overcome the subjective nature of a test and get a fairer score is to combine an oral test with

objective tests such as reading or listening comprehension and a grammar test (Ellis and Johnson, 2000).

CHAPTER 3

THE STUDY

3. 1. The context of the study

The participants of this study were six small groups of in-company intermediate students and four teachers who work in private companies in Buenos Aires.

The number of students in these groups ranged from 2 to 5. Each group had two weekly clock hours of in company English lessons. The test takers were, according to Weigle's classification of groups, adults learning a foreign language for personal interest or career enhancement (Weigle, 2002).

All the participants in this study sit for a placement exam when the year starts and an achievement test at the end of the year which is given by an external institution. The selected learners were those who got an average of 450 to 500 points in the TOEIC test or a 50% in a placement test provided by an external teacher in March, 2010.

These in-company students are considered to be mid-intermediate since considering Brown's (2000) classification they are able to satisfy most work requirements with language usage that is often, but not always, acceptable and effective.

The *ACTFL Proficiency Guidelines* (1986) which are related to the ILR (Interagency Language Roundtable) attempt to judge the test takers' level of proficiency out of eleven suggested levels (Brown, 2000).

The levels are the following:

- 0 Unable to function in the spoken language.
- 0+ Able to satisfy immediate needs using rehearse utterances.
- 1 Able to satisfy minimum courtesy requirements and maintain very simple face-to-face conversations on familiar topics.
- 1+ Able to initiate and maintain predictable face-to-face conversations and satisfy limited social demands.
- 2 Able to satisfy routine social demands and limited work requirements.
- 2+ Able to satisfy most work requirements with language usage that is often, but not always, acceptable and effective.
- 3 Able to speak the language with sufficient structural accuracy and vocabulary to participate effectively in most formal and informal conversations.
- 3+ Often able to use the language to satisfy professional needs in sophisticated and demanding tasks.

- 4 Able to use the language fluently and accurately.
- 4+ Speaking proficiency is superior in all respects.
- 5 Speaking proficiency is functionally equivalent to that of a highly articulated and well educated native speaker (Brown, 2000:112).

According to Brown (2000), the intermediate level is characterized by the students' ability to combine and recombine learned elements, to initiate, sustain and close a basic communicative task and to ask and answer questions.

Within the intermediate level, Brown (2000) describes the following differences: Low-intermediate students are those who can handle a limited number of interactive situations and are able to express the most elementary needs even if a strong interference from native language may occur. Mid-intermediate learners can handle successfully a variety of basic communicative tasks even if their speech may present long pauses due to the struggle to produce appropriate language forms (Brown, 2000). At a high-intermediate level, students are able to handle most uncomplicated communicative tasks and use a number of strategies appropriate to a range of topics. However, errors are evident at this level as well as limited vocabulary in certain situations (Brown 2000).

3. 2. Procedure and research methods

Throughout 2010, a careful selection of participants was made by the teachers who participated in this investigation after they received the results that students got from the placement test given by an external institution. Once the population was determined, appointments for the interviews were made with every teacher in charge of the groups under study and a date was agreed to test their students for one hour.

In order to collect as much data as possible, three different instruments were designed for the present investigation: an interview for teachers, a questionnaire for students and a grammar test. The aim of using a semi-structured interview was to generate quantitative and qualitative data because as Maykut and Morehouse clearly state: “the data of qualitative inquiry is most often people’s words and actions, and thus requires methods that allow the researcher to capture language and behaviour” (1994:46). Another instrument used for data collection was a self-administered questionnaire which was given to the students who participated in this study together with a grammar test which served as another research method used for content analysis in the hope of identifying specific characteristics. (De Marrais & Lapan, 2004).

In trying to verify or refute the hypotheses about students’ test results, a test was given to the participants during one of their lessons. The test consisted of

four tasks which tested the same grammatical structures with the only difference that two were closed-items and two open activities.

Once the students had completed the test, they were given a questionnaire which consisted of five questions aimed at finding out their perceptions regarding the type of exercises used for testing and eight short questions about their educational background. This questionnaire was carried out in Spanish to avoid lack of comprehension.

Another data-collection method consisted of a written interview to the four teachers who were in charge of the seventeen students tested in an attempt to compare the teachers' and students' attitude to exam content.

The data were analysed both qualitatively and quantitatively and the findings were dealt with separately but were concluded generally in an attempt to bring light to the issue under study.

Methods

3. 3. a. Grammar test

The written grammar tests tasks given to each group of students in one of their English lessons aimed at evaluating the participants of this investigation in the hope of finding out whether the type of test item used influenced their results.

This instrument was carefully designed following the requirements in Alderson, Clapham and Wall's *Language Test Construction and Evaluation*.

The participants of this research were tested on grammar since it is an area of language learning which can be tested more objectively than the reading, listening, writing and oral skills (Alderson, 2005); besides as Hughes (2003: 143) states, "Grammatical ability, or the lack of it, sets limits to what can be achieved in the way of skills performance. The successful writing of academic assignments, for example, must depend to some extent on command of more than the most elementary grammatical structures."

The same grammatical structures were tested on each of the written exercises so that the results could be analyzed in terms of the difficulty of the task rather than the grammatical structure itself. In addition, the same structures were tested throughout to reduce the element of chance, which is highly criticized, especially, in close-item tests. Even though the test included different formats,

the context provided was different so that the variable of an unfamiliar context would be reduced.

The test consisted of four parts. The first part included eight multiple choice items within the same context and each item presented four alternatives and only one correct answer. The fourth exercise consisted of closed items which attempted to test the same structures in isolated sentences and not in a given context.

The second exercise was a fill-in-the-blanks activity. The blanks were not chosen at random but rather selected carefully to check whether the same structures tested in the previous parts could be produced by the students instead of being merely recognized. Last but not least, an integrative task was included in which students were requested to write an email in response to the prohibition of the use of the internet in the company. The purpose of this task was to check if students were able to produce the grammatical structures tested in the other written items correctly.

Thus, two multiple choice exercises were included where students were asked to recognize rather than produce a certain structure. One of these items was designed in a context and the other was constructed in isolated sentences. The other two tasks dealt with the same grammatical structures as the close items

but required a thorough integration and production of the students' skills since no alternatives were offered to facilitate the task.

Item 1: Multiple choice in context

The following multiple choice item was included in the test in an attempt to find out students' results when they were given alternatives to fill in a text.

Many companies 1-_____ to make money from the Internet in these last years but it seems that people did not buy enough to keep the dot.com retailers in business. Advertisers 2-_____ pay to get their products at the top of the list when people search for information. Sixty thousand advertisers, for example, use Google's services 3-_____ worldwide. This is the world's 4-_____ search engine 5-_____ has AOL and Earthlink as its clients. If Google is 5- interested in staying in first place, it will have to keep prices low. The Managing Director explained that advertisers give the company 6-_____ profit so they don't make enough money from them. But as there is strong competition these days, the company 7-_____ to offer a good service. Therefore, Google is trying to provide its customers with services 8-_____ are different from their competitors.

1- A- try	B- has tried	C- have tried	D- is trying
2- A- ought to	B- should	C- have to	D- has to
3- A- effective	B- more effective	C- effectively	D- more effectively
4- A-bigger	B- big	C- biggest	D- the biggest
5- A- where	B- who	C- where	D- which
6- A- much	B- few	C- many	D- little
7- A- work	B- had worked	C- worked	D- is working
8- A- who	B- whose	C- which	D- where

The purpose of including four alternatives was not only to reduce the element of chance but also because alternatives such as 2a and 2b or 3a and 6a seem grammatically correct but are not appropriate in the given context.

Item 2: Gap-filling

The second part of the test consisted of a paragraph with blanks to complete but no alternatives were given and not any word fitted. The context, along with the students' knowledge of grammar, should be enough to fill in the task. Also, the same structures used in the item above were repeated in this exercise.

Andrew Miwa, China's Porsche salesman, is ready for the _____ important boom in the car business since 1920. In Beijing alone, about 4,000 new cars arrive every week and 600,000 new drivers _____ get licenses next year. However, Mr. Miwa, _____ goes to work by car every day, has realized that the number of cars _____ causing great problems. The government has _____ destroy many areas to change roads into highways to avoid so _____ traffic. "People love going _____ work by car because it is _____ but something should be done" Miwa says.

It was the purpose of this closed item to examine whether test takers were able to fish for the appropriate word even though the only guideline was the context in which the word appeared and no alternatives were presented.

According to Oller (1989), gap-filling items measured the same type of skills as productive tests since they require the learners' ability to integrate grammatical, lexical, contextual, and pragmatic knowledge in order to supply the missing words. Therefore, this type of item will be considered an open task for the sake of this study.

Item 3: Writing

Task three was aimed at discovering what learners could produce when an open item is included in a test in an attempt to encourage the use of the structures acquired.

Your company has decided that the use of the Internet should be prohibited during working hours. Write an e-mail to a friend telling him why the company has come to this decision and describe how the situation has changed.

This task sought to trigger the use of modal verbs by asking test takers to write whether they agree or not with the company's new policy. Also, students were asked to describe why the company has made this decision which fosters the use of tenses. The description of how the situation has changed attempted to invite the participants to compare their jobs when the use of the internet was allowed and when it was forbidden. The use of modal verbs, tenses, comparatives and superlatives are also tested in the other test items.

Item 4: Multiple choice in isolated sentences

The last exercise was designed in order to collect data on students' grammar performance of the same structures tested above but in a closed item where each sentence was presented in isolation.

- 1- Julia English for ten years.
 a- have studied b- studies c- is studying d- has studied
- 2- Steve Jobs, is the owner of Apple, is about to retire.
 a- whom b- who c- that d- whose
- 3- Katherine arrived early because she had too..... work.
 a- few b- little c- much d-many
- 4- I've got three brothers. The lives in Europe.
 a- elder b- older c- elderly d- eldest
- 5- Tim Burton is my uncle. He is a lawyer.
 a- success b- successful c- successfully d- more successful
- 6- There is a lot of food for dinner so you bring anything.
 a- shouldn't b- mustn't c- can't d- don't have to
- 7- Paul in the Finance Department since he graduated.
 a- has worked b- had worked c- have worked d- works
- 8- My boss a project at the moment to set up a new business abroad.
 a- makes b- is making c- made d- has made

The purpose of including this closed item in the test was to find out whether students were able to complete the isolated sentences by choosing the correct word out of the options given.

3. 3. b. Analysis of each test items

Each type of written item in this grammar test intended to collect data on test takers' performance on the grammar contents listed below.

Verb Tenses

The following items requested students to complete the gaps with appropriate verbs tenses in order to provide insight into their knowledge of the present perfect, continuous and the future.

Item 1

1- Many companies *have tried* to make money from the Internet in these last years.

7- But as there is strong competition these days, the company *is working* to offer a good service

Item 2

Mr. Miwa who goes to work by car every day has realized that the number of cars *is causing* great problems.

In Beijing alone, about 4,000 new cars arrive every week and 600,000 new drivers *will get* licenses this year.

Item 3

When the students were asked to explain why the company had come to the decision of forbidding the use of the internet or how the company had changed. The aim was to elicit the use of the present perfect and also the present continuous to describe the way they are working at the moment along with the simple past, if necessary.

Item 4

d- Julia *has studied* English for ten years.

e- Paul *has worked* in the Finance Department since he graduated.

f- My boss *is making* a project at the moment to set up a new business abroad.

Modal Verbs

These sentences were designed as way of checking to what an extent students could handle the use and meaning of verbs expressing modality.

Item 1

Advertisers 2- *have to* pay to get their products at the top of the list when people search for information

Item 2

The government *has* to destroy many areas to change roads into highways.

Item 3

The use of *can't*, *mustn't* or *should* was expected when describing what the situation was like after the internet was forbidden.

Item 4

There is a lot of food for dinner so you *don't have to* bring anything.

In all the items selected, the concept of obligation and lack of obligation was tested to find out whether the students were able to recognize or produce modality in spite of the context in which it appears or the format in which it is presented.

Word categories

Choosing the appropriate word category might offer some difficulty especially when students should fit it in a given sentence. The purpose of the items below was to examine how familiar students were with the use of adjectives and adverbs.

Item 1

Sixty thousand advertisers, for example, use Google's services *3-effectively* worldwide.

Item 2

People love going to work by car because it is *comfortable*, but something should be done.

Item 3

Participants were expected to use either adjectives or adverbs to express how their jobs had changed after the decision the company had made.

Item 4

Tim Burton is my uncle. He is a *successful* lawyer.

Comparative and Superlative forms of Adjectives

To a certain extent, this item overlaps with the item mentioned above since most distractors included in the alternatives given belonged to this grammar structure. A further analysis might be useful when comparing the results that students obtained.

Item 1

This is the world's 4- *biggest* search engine.

Item 2

Andrew Miwa, China's Porsche salesman, is ready for the *most* important boom in the car business

Item 3

Compare how the situation has changed.

Expected sample answer: People will finish work earlier and will work faster than before.

Item 4

4- I've got three brothers. The eldest lives in Europe.

In all the items, except for the open task, the superlative form of the given adjective was required and the explicit use of "the" paved the way to the right word. Item two intended to encourage students to produce the superlative form of two or more syllable adjectives.

Relative Clauses

Relative clauses were also part of the test and they were included in four sentences.

Item 1

5- This is the world's biggest search engine 5- *which* has AOL and Earthlink as its clients.

8- Google is trying to provide its customers with services 8- *which* are different from their competitors.

Item 2

However, Mr. Miwa, who goes to work by car every day, has realized that the number of cars is causing great problems.

Item 4

2- Steve Jobs, *who* is the owner of Apple, is about to retire.

The use of “who” and “which” is tested throughout in an attempt to find out if test takers guessed the right option, if they were hindered by the context in which the word appeared, if the lack of context prevented them from choosing the right alternative or if they were really able to tackle the structure.

Countable and Uncountable Nouns

Countable and uncountable nouns were introduced in each item through sentences which required the use of “much” or “many”, “few” and “little”.

Item 1

The Managing Director explained that advertisers give the company *6-little* profit so they don't make enough money from them.

Item 2

The government has to destroy many areas to change roads into highways to avoid so *much* traffic.

Item 4

Katherine arrived early because she had too *much* work.

The sentences included in the items above all dealt with uncountable nouns. Therefore, the use of *much* and *little* was required. The first multiple choice item presented the highest difficulty since students should understand not only the rule for uncountable nouns, but also the context that surrounds this word, which gives meaning to the paragraph.

3. 4. Questionnaire

This instrument consisted of two parts. The first section included four questions regarding test takers' perception on the type of written items used for testing their grammar knowledge.

The second group of questions aimed at finding out a little about the participants' background in the hope of collecting further data which might contribute to accounting for differences in the test results of students who have the same teacher.

It was the purpose of the first question to get to know students' experience with multiple-choice items and their attitude to the alternatives which are given in this type of exercise since although choosing out of a number of options might be easier than coming up with the right answer, the alternatives given might be confusing or challenging for those test takers who are not so familiar with the structure that is being assessed.

Question 1

1- Las alternativas que se dan en un examen de respuestas múltiples suelen ser:

- a- capciosas
- b- complejas
- c- sencillas

d- otro ¹

The aim of this question was to find out students' perception of the alternatives given in multiple choice items in order to compare their answer to their results later on.

a- The option *tricky* was used because when a task seems easy, students fear that they might be missing something.

b- *Difficult* intended to reflect how students find those alternatives that are unknown to them, common mistakes or similar to the right option and present a challenge for the student because they would not have thought of a certain option but by including it within the alternatives, doubts might arise.

c- Some students find close item much easier than performance tests because when alternatives are given, the element of chance appears or the expected answer is reduced to the options presented.

d- This alternative invited participants to make any other comment they wanted.

¹ The alternatives given in a multiple choice item are:

- a- tricky
- b- difficult
- c- easy
- d- other

Question 2

Participants were asked to rank the following items used for testing depending on their level of difficulty.

2- ¿Cuáles de los siguientes ejercicios te resulta más sencillo? Numerar del 1 al 7 (siete es el más complejo)

- a- Unir con flechas palabras, frases o respuestas.
- b- Llenar los espacios vacíos de un texto con una palabra o frase.
- c- Elegir una respuesta correcta entre varias alternativas.
- d- Responder preguntas que pueden ser verdaderas o falsas.
- e- Ordenar palabras para formar oraciones.
- f- Transformar palabras para completar correctamente un contexto dado.
- g- Identificar y corregir errores gramaticales en un texto.
- h- Escribir una composición.²

² Which of the following items do you find more difficult? Rank them from 1 to 8

- a- Pairing and Matching (matching responses from the alternatives given)
- b- Multiple Choice (selecting a response from among alternatives provided)
- c- Gap-Filling (restoring the missing word in a given passage)

It was the aim of this question to discover which of the items often included in language tests students find easy or difficult.

Items *a*, *d*, *e* and *f* were not considered in the test whose results were used for content analysis in the present study. However, they were included on the list in an attempt to compare students' ranking with the teachers' comments on the influence of the students' ability or training in their test results.

Items *b*, *c* and *h* were part of the grammar test so their rank will be later compared with the way students' performed in each of these items.

Question 3

The aim of this question was not only to double check the answers in the first and second question but also to examine to what an extent test takers found open tasks more difficult than close items and to triangulate these data with students' test performance.

d- Dichotomous Items (True/ False- Yes/ No- questions)

e- Ordering (arranging "broken" sentences)

f- Transformation (changing the structural pattern of a word/ phrase)

g- Editing (identifying the errors which have been introduced in a given text)

h- Writing

3- En la evaluación de gramática ¿Cómo te resultaron los ejercicios de respuestas múltiples?

a- más fáciles que escribir una composición.

b- más difíciles que escribir una composición.

c- igual de complejos como escribir una composición. ³

Question 4

4- ¿Cuándo te sentís más seguro en la elección de la respuesta?

a- cuando se ejercitan los temas en un determinado formato en clase.

b- hay formatos que te resultan naturalmente más sencillos.

c- depende del tema, varía la dificultad del formato. ⁴

³ In the grammar test, how did you find the multiple choice exercises?

- a- Easier than the writing task
- b- More difficult than the writing task
- c- As difficult as the writing task

⁴ When do you feel more confident about the right choice?

- a- when you are given exercises in different formats in class
- b- there are items that you find naturally easier than others
- c- the difficulty of the item depends on the difficulty of the structure under study

Test takers might sometimes be hindered either by the alternatives suggested or by the lack of options in a given test item. Also, when there is more than one correct answer, the item might become either easier to complete or more difficult, depending on how familiar the students are with the structure and the type of item. Moreover, there are cases in which students' practice in class or at home is not in keeping with the way they are assessed and this might influence their test results. Option b attempts to explore which participants think that the results depend on their abilities rather than on their practice in the hope of establishing a comparison with teachers' interview in the triangulation. The last alternative seeks to find out whether the difficulty arises from the topic that is being tested rather than on the type of item.

The second section of the questionnaire aimed at exploring the participants' background since although they all attended an intermediate course of an in-company English class, the students' experience with the language can be very different. Therefore, the following questions were included:

1-¿Cuántos años estudió ingles? <input type="checkbox"/> 1 a 2 años <input type="checkbox"/> 3 a 4 años <input type="checkbox"/> 5 a 6 años <input type="checkbox"/> mas de 6 años ⁵
--

⁵ 1- How long have you studied English?

1 to 2 years 3 to 4 years 5 to 6 years more than 6

The aim of this question was to find out if the participants shared the same level after studying English the same number of years or not. An intermediate student who has studied for more than 6 years might have different learning abilities from someone who has acquired the language in two years.

2- ¿Realizó estos estudios con continuidad?
<input type="checkbox"/> si <input type="checkbox"/> no ⁶

Sometimes it is not only the time one spent studying but also the continuity of the course. Some people have studied for many years and haven't practised for a long time and their performance might be influence by this factor.

3-¿Que tipo de cursos ha tomado?
<input type="checkbox"/> particular <input type="checkbox"/> grupales <input type="checkbox"/> ambos <input type="checkbox"/> otro ⁷

The type of language instruction a person receives individually is not the same as the learning environment in a large group. Therefore, this is another variable that was later considered for the analysis of data.

⁶ 2- Have you studied all these years non-stop?
 yes no

⁷ 3- What type of courses have you attended?
 private lessons in groups both

4-¿Donde ha estudiado?

en el colegio en un instituto en el trabajo otro⁸

Participants' performance might depend on the experience they had while learning the language since the instruction learners get is very different when one attends a course a couple of hours in a school of English, when one is exposed to the language in school at an early age or when one studies in company for business purposes.

5- ¿Qué tipo de exámenes ha rendido?

de nivel de fin de curso internacionales otro⁹

Success in test results might depend on previous training. Students' experience with different type of test might be an advantage. The aim of this question is to learn about participants' experience with testing.

6- ¿Con cuánta gente ha compartido las clases de inglés?

menos de 5 entre 5 y 10 menos de 20 mas de 20¹⁰

⁸ 4- Where have you studied?

in school in an institute in company

⁹

5- What type of exams have you sat for?

placement test achievement test international exams

¹⁰ 6- How many students shared the lessons with you?

This question intends to reveal whether the “group” that was considered in question three was a large or small course.

7- ¿En qué contextos utiliza ud el inglés? <input type="checkbox"/> solo en clase <input type="checkbox"/> en el trabajo <input type="checkbox"/> viajes de negocios <input type="checkbox"/> otro ¹¹

This question intends to check how authentic the use of the language is for each participant. Students who use English for business purposes might develop communicative strategies that students who only use it in the learning environment lack.

This questionnaire was aimed at identifying and describing the groups of students and teachers that participated in the study in order to facilitate the triangulation of the instruments of the investigation.

less than 5 between 5 and 10 less than 20 more than 20

11

7- When do you use the language?

only in class at work business trips

The items considered were those that might explain an important difference in the performance of students of the same group or if there is incongruence between their results or perceptions and their teachers' experience.

Although all participants belonged to an intermediate group, the number of years studying the language, the type of courses they attend, their previous experience with other exams and the purpose of studying the language might influence their performance.

3. 5. Semi-structured interview

This instrument consisted of six questions which intended to collect data about teachers' experience and perceptions in testing their students' grammar performance.

In questions one and two teachers are asked to choose alternatives from a list of written items used for testing. Questions three, four and six require a more extensive response on the part of the teacher regarding their experience; in the fifth question; teachers are requested to rank the test items depending on their level of difficulty.

Question 1

1- Which of the items listed below do you use in class?

- Pairing and Matching (matching responses from the alternatives given)
- Multiple Choice (selecting a response from among alternatives provided)
- Gap-Filling (restoring the missing word in a given passage)
- Dichotomous Items (True/ False- Yes/ No- questions)
- Ordering (arranging “broken” sentences)
- Transformation (changing the structural pattern of a word/ phrase)
- Editing (identifying the errors which have been introduced in a given text)
- Writing

This question was aimed at identifying the type of exercises that the teachers who participated in this study include in their tests in order to facilitate triangulation with other instruments.

There are several terms to refer to fixed and open response tasks. In this interview, McNamara’s (2007) classification was used which also includes Oller’s and Alderson’s contribution to defining the different formats which can be included in a test. A short description of each item was included to help teachers identify the types of tasks.

Question 2

2- Which of these items do you include when you test your students' knowledge of grammar?

- Pairing and Matching (matching responses from the alternatives given)
- Multiple Choice (selecting a response from among alternatives provided)
- Gap-Filling (restoring the missing word in a given passage)
- Dichotomous Items (True/ False- Yes/ No- questions)
- Ordering (arranging "broken" sentences)
- Transformation (changing the structural pattern of a word/ phrase)
- Editing (identifying the errors which have been introduced in a given text)
- Writing

The purpose of this question was to double check whether the items chosen in question one were included and to learn what formats are specifically used for testing grammar since the focus of this investigation was on students' grammar performance. A further objective to this question was to collect data to later compare with the students' results in the test.

Items a, b, c and h were included in the grammar test constructed as a useful instrument. Practical reasons and time limitations account for the omission of

items d, e, f and g in the test. However, they were included in this question for further analysis and discussion.

Question 3

3 – Why do you prefer these items? (They are easier to grade, they are objective, they also test comprehension, other learning strategies are tested etc.)

This part of the interview sought to register comments on teachers' experience with test items. Some choices were listed not only to elicit as much information as possible but also to find out whether teachers used certain formats out of personal choice or because tests were designed by other authorities.

Question 4

4- In your experience, which of these items offer greater difficulties? Why?

In an attempt to gain insight into the teachers' perception of their students' performance, question number four was included. The objective was to identify which items students find more difficult. The ulterior purpose of this section was first, to triangulate teachers' responses, students' comments and their test results and then, to examine whether teachers believe that students fail because they do not know a certain grammatical structure or because the format of the task is too hard for them to complete successfully.

Question 5

5- Which, in your opinion, reflect better your students' knowledge of grammar?

- Pairing and Matching (matching responses from the alternatives given)
- Multiple Choice (selecting a response from among alternatives provided)
- Gap-Filling (restoring the missing word in a given passage)
- Dichotomous Items (True/ False- Yes/ No- questions)
- Ordering (arranging "broken" sentences)
- Transformation (changing the structural pattern of a word/ phrase)
- Editing (identifying the errors which have been introduced in a given text)
- Writing

It was the aim of this question to focus on teachers' opinion on the type of exercises used to test grammar and also, to find out their main purpose when designing a test since some items might be useful for testing objectively but difficult for testing grammar performance.

Question 6

6- Do you think the items mentioned above require previous training? Why?

The objective of this last question was to give teachers the opportunity to reflect on their own choices since sometimes there is incoherence in the type of practice students get and the way they are tested. Therefore, this question aimed at discovering how teachers go about the teaching and testing process.

CHAPTER 4

THE RESULTS OF THE STUDY

1- Grammar test

As has already been pointed out, a grammar test which included four different written items was given to six groups of intermediate students of English attending in-company lessons.

For the sake of this study, the groups which participated in this investigation will be named with the letters of the alphabet and the students will be identified with numbers. The participants of each group and the teachers in charge of the course are listed below:

Group A Teacher I Students 1, 2 and 3

Group BTeacher II Students 4, 5 and 6

Group C Teacher II..... Students 7 and 8

Group D Teacher IIIStudents 9, 10, 11 and 12

Group ETeacher IIIStudents 13 and 14

Group F Teacher IV..... Students 15, 16 and 17

The tests were included in the appendix.

Group A

Student 1

This student managed to complete most exercises correctly.

Item 1: In the paragraph designed to complete with one of the given alternatives, the test taker chose all the right options except for the first one, which was testing the students' knowledge of the present perfect.

Item 2: the learner succeeded in filling the blanks correctly except for the gap where the word "many" was written instead of "much", which shows that he knew the type of word that fitted but failed to identify the influence of the uncountable noun "traffic".

Item 3: Verb tenses were properly used throughout the text and although most of the structures tested in the other items were omitted, most errors were spelling mistakes.

Item 4: The same mistake made in item 1 was repeated in the last item, which also consisted of multiple choices but in isolated sentences. This student chose to use the present continuous in both sentences where the present perfect was expected.

Student 2

What is interesting to point out about the results of this test is that this student completed only one verb tense item correctly out of six that appeared throughout the exam. However, in the open item task where students were invited to write freely about a given topic, all the verb tenses included were used appropriately.

Item 1: there were 3 mistakes out of 8 gaps to complete.

Item 2: the gap filling item which did not provide any alternatives was successfully completed except for the item where the relative pronoun “who” was more suitable than “he”.

Item 3: the open item task did not show any meaningful mistakes.

Item 4: the only wrong answers were the choice of a comparative instead of a superlative adjective and the recognition of tenses.

Student 3

Item 1: two out of the three mistakes in this part referred to verb tenses and the other mistake was on the choice of superlative adjectives.

Item 2: this item was well completed except for the gaps testing the future form “will” and the relative pronoun “who”.

Item 3: this item has not been completed.

Item 4: only one mistake was made in this section since “many” was selected instead of “much”.

Group B

Student 4

This student got most of his mistakes in the first and second test item.

Item 1: the two gaps testing relative pronouns were the only correct answers chosen in this item.

Item 2: surprisingly, only two mistakes were made in this section. One of these was the use of “have to” and the other was the use of “he” where the relative pronoun “who” was expected.

Item 3: the test taker managed to pass on a message through the use of nouns, verbs and adjectives. However, the structure of the text is very poor and none of the expected grammar items were used.

Item 4: not even relative pronouns which were used correctly in item 1 were recognized in this test item. The only right answers are the ones that tested the use of tenses.

Student 5

Item 1: there was only one mistake in the choice of a comparative adjective instead of the superlative.

Item 2: this was the weakest part of the exam. However, the structures which were not written in this item were properly used in the other activities.

Item 3: the use of verb tenses, uncountable nouns and comparatives was included and even though there are some language mistakes, these do not hinder comprehension.

Item 4: all the sentences were properly completed.

Student 6

Item 1: the answers chosen for tenses, modals and word categories were incorrect.

Item 2: the same grammar mistakes made in exercise 1 were made in this item along with the omission of the relative pronoun “who”.

Item 3: In spite of using very simple language and structures in the open task, this participant managed to write a clear and grammatically correct response.

Item 4: The same mistakes that were made in item 1 and 2 along with the misuse of countable and uncountable nouns appeared in this last item

Group C

Student 7

Unlike the other groups, this student presented more mistakes in the last item and made a very good performance in the first three activities.

Item 1: all the alternatives chosen were correct.

Item 2: all the gaps were properly completed except for the use of the relative pronoun “who” and the use of “many” instead of “much”.

Item 3: this participant succeeded in producing a clear and accurate paragraph using the expected grammar points.

Item 4: two mistakes were made in this section and both tested verb tenses.

Student 8

The right choice was selected in almost all the items.

Item 1: all the answers given were correct except for the identification of uncountable nouns.

Item 2: all the gaps were properly completed.

Item 3: many grammar mistakes were made in this item but none of these errors prevented the participant from passing on a clear message.

Item 4: all the alternatives chosen were correct.

Group D

Student 9

Item 1: there was only one mistake in the choice of grammatical categories since the phrase “more effective” was selected instead of “effectively”.

Item 2: only three gaps were completed and none of them was appropriate.

Item 3: this was the weakest part of the exam since the few sentences that were written were ungrammatical and the overall idea was not clearly stated.

Item 4: all the sentences were properly completed.

Student 10

Item 1: all the alternatives chosen by this student were correct except for two sentences which tested the use of verb tenses.

Item 2: all the gaps were properly completed.

Item 3: the open task did not present any mistakes regarding tenses.

Item 4: all the alternatives chosen were correct.

Student 11

Item 1: one mistake in this section was regarding the recognition of countable and uncountable nouns, another in the choice of an adjective instead of an adverb and the last error was in the choice of the present perfect.

Item 2: all the gaps were properly completed except for the use of “many” instead of “much” in one sentence, which is the same mistake made in item 1.

Item 3: this student failed to use the comparative and some verb tenses appropriately. However, he/ she managed to pass on a clear message.

Item 4: the mistakes made on item 1 and 2 were not repeated in the last item.

Student 12

Item 1: two out of the three mistakes in this test referred to verb tenses and the third error was in the identification of grammatical categories.

Item 2: this item was completed successfully except for the gap testing the present continuous where the auxiliary “are” was written instead of “is” and the gap where the relative pronoun “who” was expected was left incomplete.

Item 3: there were attempts to use the present perfect but this test taker failed to write the main verb in the past participle form. Although there are other mistakes, ideas are clearly stated throughout

Item 4: Identifying the present perfect and the superlative adjective were the only mistakes in this activity.

Group E

Student 13

The student's grammar weaknesses were clearly shown in this test since the same mistakes were repeated throughout in spite of the different context and formats which distinguish one exercise from the other.

Item 1: verb tenses, grammatical categories and the identification of uncountable nouns presented difficulties.

Item 2: the same structures misused in item 1 were left incomplete in this activity.

Item 3: the open task was well organized and there were no spelling mistakes. However, most verb tenses were used incorrectly.

Item 4: only the sentence that tested the use of the present perfect was completed incorrectly.

Student 14

Items 1, 2 and 4 were completed successfully except for one minor spelling mistake in the gap filling exercise. Although all the test takers were given enough time to make the test, this student did not produce any piece of writing in item 3.

Group F

Student 15

Item 1: only two gaps were properly filled.

Item 2: one of the grammar structures which was chosen correctly in item 1 was left incomplete in the second item.

Item 3: the open task was carried out with very simple language but ideas were clearly expressed.

Item 4: there were only three gaps properly completed. None of them tested the structures which were well used in item 1 and 2.

Student 16

Item 1: only two mistakes were made by this student and both appeared in this item. One of the mistakes was in the choice of a comparative adjective instead of an adverb and the other was in verb tenses.

Item 3: this participant's writing performance was really good.

Item 2: this part was completed successfully.

Item 4: there were no mistakes spotted in this item.

Student 17

Item 1: two mistakes were found in the sentences which tested the recognition of verb tenses.

Item 2: all the gaps were properly completed except for one minor spelling mistake.

Item 3: this participant succeeded in producing a clear and accurate paragraph using the expected grammar points.

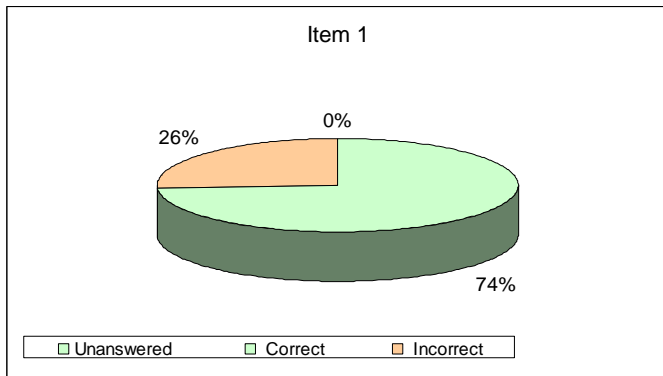
Item 4: no mistakes were made in this section.

Test items

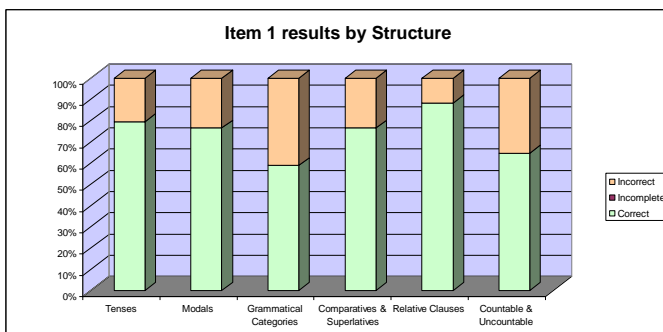
The data listed above can be analysed in terms of the grammar structure that is being tested or it can be studied according to the results obtained by students and their teachers' beliefs on testing. The aim of this study is to focus on the results obtained in each test item in order to draw global findings about the participants' performance in different types of test tasks.

Item 1

This first item was designed as a multiple choice exercise which consisted of a short paragraph where eight gaps had to be completed with one right answer selected out of four alternatives each. In spite of being presented with four choices in each sentence, which might generate doubts as regards the most suitable word to fill in the gap, 74% of the answers chosen by the seventeen test takers were correct.



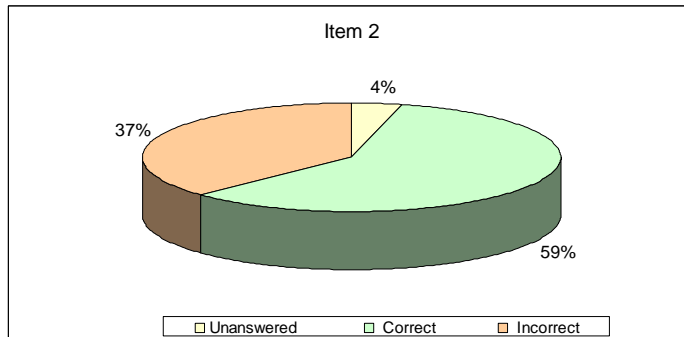
Within the first item, more than one grammar structure was tested. The gaps which required the use of relative clauses were completed correctly by 90% of the participants and the selection of word categories such as the use of adjectives and adverbs, showed the lowest number of correct responses. All the blanks which were included in this test item were completed.



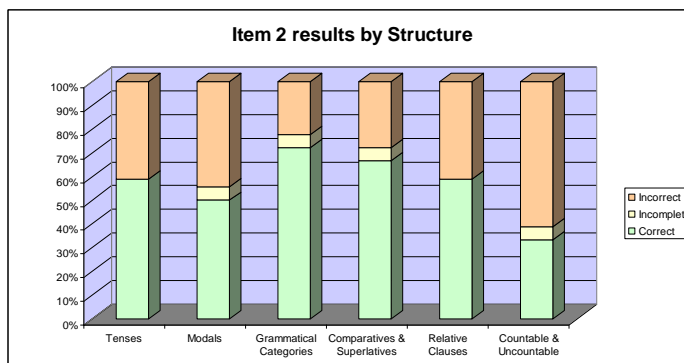
Item 2

Students' results were also analyzed in item 2 which consisted of a gap-filling exercise. Unlike item 1, this activity required students to resort to their

knowledge of grammar to produce the words which best fitted the blanks for no alternatives were given. The outcome is shown in the following chart.



In this test item, 4% of the blanks were left incomplete and 59% were filled correctly. In order to carry out this task successfully, participants needed both comprehension of the paragraph designed and enough knowledge of grammar to produce the expected response.

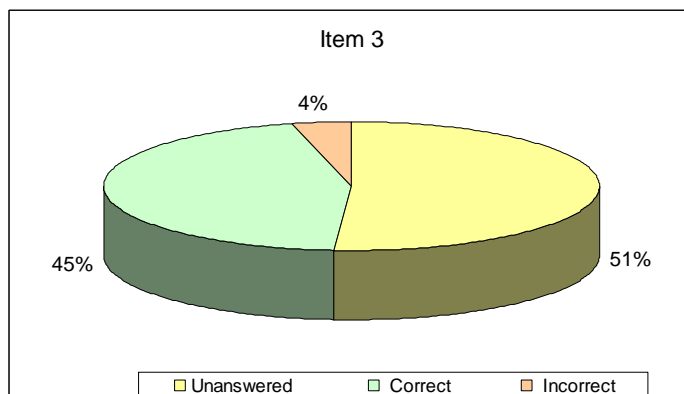


There were structures such as relative clauses which were chosen correctly in item 1 but which almost half of the students failed to produce appropriately in this item. The proper use of countable or uncountable nouns in this section

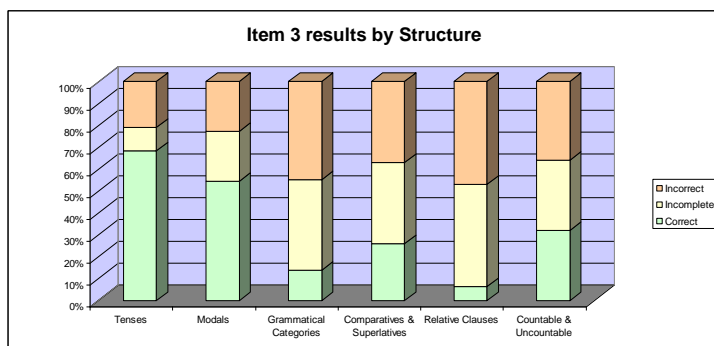
brought difficulties too since less than 40% managed to find the right answer. However, if compared to item 1, the percentage of wrong answers was not so high considering that no options were given and any word could fit the gaps.

Item 3

This was an open task where students were asked to write an email expressing ideas on a given situation. Surprisingly enough, 45% of the implicitly expected structures were properly used. However, as seems to be the case of most open tasks, 51% of these structures were omitted and in most cases they were replaced by simpler sentences.



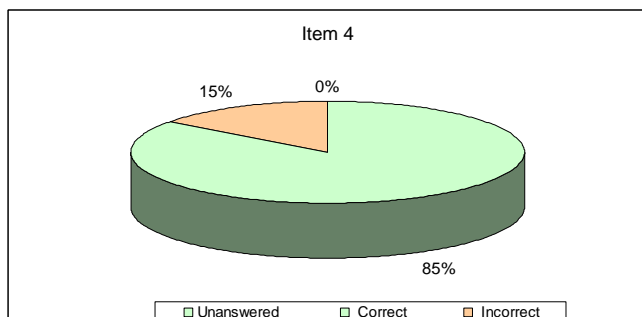
The use of verb tenses and modal verbs was correct in over 50% of the cases. In contrast to item 1 but similarly to item 2, relative clauses were mainly incorrect, if used.



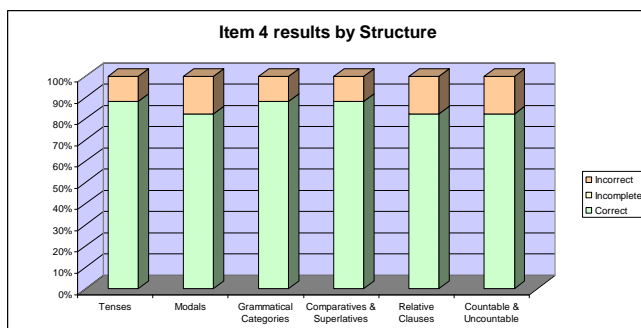
Item 4

Four alternatives were presented to complete the gaps in isolated sentences.

This item was included especially because it is the format used by the external institutions that assess the participants of this study.



Only 15% of the sentences in this item were completed with the wrong answer and none of the blanks were left unanswered. Even structures that were misused in item 1, omitted in item 2 and avoided in item 3 were well chosen in this part.



All in all, the findings of this instrument have shown that better results were obtained in items 1 and 4. Both consisted of a task with options out of which the correct answer had to be identified. However, the outcome of the first item, which presented gaps in sentences within one paragraph, showed a lower percentage of correct answers than the multiple choice sentences designed in isolation.

As regards item 2, which entailed more than just selecting the right choice, participants managed to complete only 59% of the blanks successfully even though most of the words expected in these gaps appeared somehow in the first item. Likewise, the results obtained in item 3, which required students' writing performance, have shown that the production of certain grammar points is more challenging than their recognition. 45% of the expected structures were properly used and although half of them were completely omitted, only 4% were incorrectly adopted.

Focusing not only on the results obtained in each exercise but also on the structures tested, it would be worth noticing that the percentages of correct answers in the four test items which assessed relative clauses showed the biggest difference. The outcome of items 1 and 4 was higher than 80%, whereas the average of correct responses in item 2 was of only 59% and in the writing task the lowest percent of all was calculated with a 12%.

Although the results shown in the use of comparative and superlative nouns were not as low as that of relative clauses, it was surprising to see the difference in the outcome of each item. Testing this structure in different formats seemed to have had some kind of influence in the results since item 1 showed 65% of correct answers, 35% were obtained in item 2, almost 50% in the writing task and 82% in the multiple choice item.

The smallest difference in the students' performance considering the four test items was registered in the use of verb tenses with an average of 76% right responses in all the items.

2- Questionnaire

The questionnaire was given to each of the participants after the grammar test in an attempt to explore students' experience with different test items.

Question 1

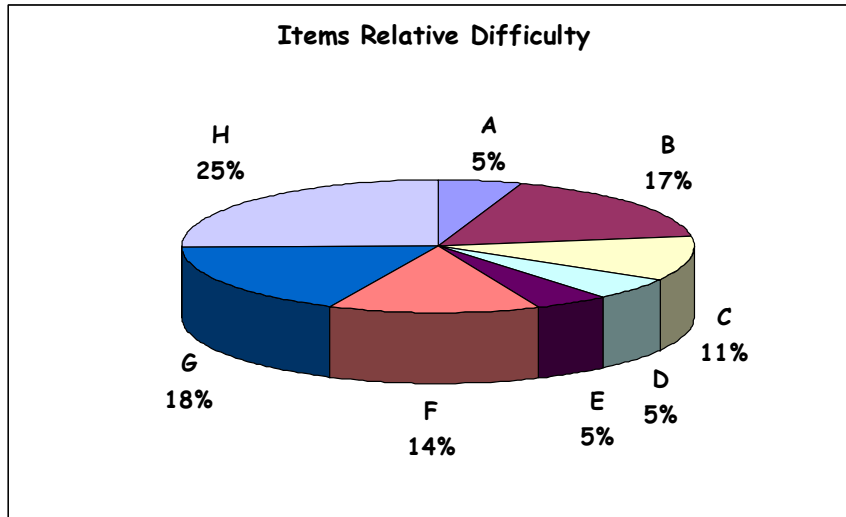
This question aimed at finding out students' perception of the alternatives given in closed-item tests.

Only one of the participants of this study considers multiple choice items easy and 20% finds them difficult. Most students agree that when they are given a closed-item with multiple choices out of which they have to select one response, the alternatives included are usually tricky or confusing. Four participants explained that there are sometimes words included in the options that they would have never thought of but when they are exposed to alternatives which are unfamiliar to them or which are not frequently used by them, doubts arise and this makes the correct choice harder to spot.

Question 2

The aim of this question was to explore students' insight about different types of test items.

Test takers were also asked to rank different type of test items in an attempt to find out to what extent their results are in keeping with their perceptions and experience.



There were eight type of items listed and only some were included in the test. The purpose of listing all these items was to make a comparison between the way participants feel about the items included in the test, the ones that were excluded and the types that they are not familiar with.

The writing task, as the chart shows, was chosen as the most difficult item regardless of the results obtained by the students. 11% of the participants chose multiple choice items as the most challenging activity and a higher percentage of students (17%) chose gap-filling. The activities included in the test used as an instrument for this study were within the ones chosen as the

most difficult except for word transformation, which turned out to be chosen by 18% of the students. The knowledge needed for this type of item was somehow included when grammatical categories were assessed and a correlation between students' perception and their results on the test will be analyzed in the triangulation of the research instruments later on.

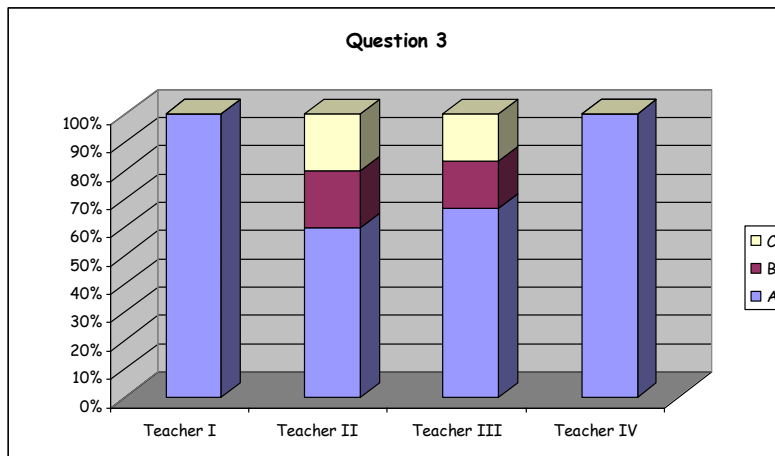
The test items which were ranked as the easiest activities by the participants of this study are pairing and matching (5%), ordering (5%) and dichotomous items (5%). Two students made a brief comment below their rank explaining that there are usually more chances of getting the right answer in pairing, matching and dichotomous items even when you are not sure about the correct response.

Editing, an item in which students are asked to identify the mistakes introduced in a text, was considered difficult by two participants and it would be interesting to point out that this is an item which is frequently included in the final assessments given to these groups.

Question 3

It was the aim of this question to learn about students' perception of the grammar test.

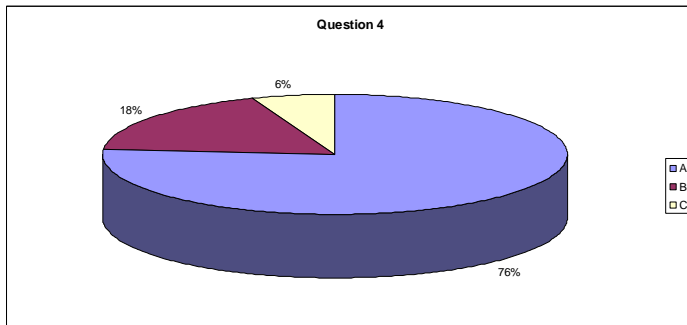
All the students in groups A and F considered the close items in the test easier than the writing task. 20% of the students in groups B and C, who share the same teacher, believe that the closed response format was more difficult than writing a paragraph and 17% of the participants who are divided into two groups found the open task as difficult as the multiple choice items. Once the teachers' information is collected, the influence of the students' results and perceptions will be compared and analyzed further.



Question 4

The purpose of including this question was to find out about students' experience in test items.

The graphic representation below shows the results obtained in this question.



The chart above describes students' answers to the way they feel about their performance regarding the training they receive, their own abilities and the topic which is being assessed.

76% of the people under study agree that previous training is necessary to get better results. The presentation of the structures which will be assessed in different formats and further practice of these items are also important for most of the students.

18% believe that it is not a matter of previous practice but rather an internal ability that some people have for certain activities. Some students are good at expressing themselves accurately in a written open task while others have better skills for completing fixed response items.

Only 6% of the participants think that the topic which is being tested might cause difficulties and therefore influence students' results in a test.

The questionnaire revealed mainly students' perceptions towards test items. The global findings were that even though multiple choice items often present confusing alternatives, they are easier to complete than open item tasks. Gap filling and word formation activities were ranked as the most difficult ones and most test takers believe that completing test items successfully requires previous training rather than a natural ability.

The second part of the questionnaire aimed at finding out about the participants' background education to analyze the variables in case big differences were found in the results of one same group.

Group A

Student 1

This student has studied English for five years in small groups of in company lessons and uses English in class and at work.

Student 2

This student has taken in-company English lessons for four years and practises English daily for business needs.

Student 3

This participant studied English for four years in a large group in school and one year at work and often travels abroad.

Group B

Student 4

English was a subject in school for three years for this participant and 2010 was his first year in an in-company class.

Student 5

This student has studied English for five years in small groups of in company lessons and uses English in class and at work.

Student 6

This student acquired his knowledge after traveling on business for many years and this was his second year under a formal instruction.

Group C

Student 7

This test taker took private lessons for almost three years and studied one year in company in a small group and needs to handle the language for work.

Student 8

This student has taken in company English lessons for five years and practices English daily for business needs and for future business trips.

Group D

Student 9

This participant devoted more than six years to studying English both in small and large groups but has not attended these courses non-stop.

Student 10

This learner studied for more than six years but in different institutions, in large and small groups, for business and for personal aims.

Student 11

In spite of studying English for only three years, this student has been in contact with English speaking customers and contexts for many years.

Student 12

The English knowledge acquired by this participant was mainly in primary school, one year in-company and only practises the language in class.

Group E

Student 13

Four years of in-company lessons are this participant's experience with English learning and test taking.

Student 14

This student's background consists of two years of private lessons in company and two years in a small group along with the daily experience acquired at work with foreign customers.

Student 15

His knowledge of English was mainly acquired through self instruction and two years of in-company lessons.

Student 16

This participant took private lessons for one year and studied two years in company in a small group and needs the language for work.

Student 17

Three years in a large course in university, one with a private teacher and eight months in company are this test takers' experience in learning the language.

3- Teachers' interview

As the teachers who participated in this study were asked both open and closed-ended question, the outcome of the interviews will be presented quantitatively and qualitatively in the hope of establishing similarities or differences with their students' results and perceptions later on.

Question 1

Teachers were given a list of items and they were asked to select the ones that they include in the tests they design for their students.

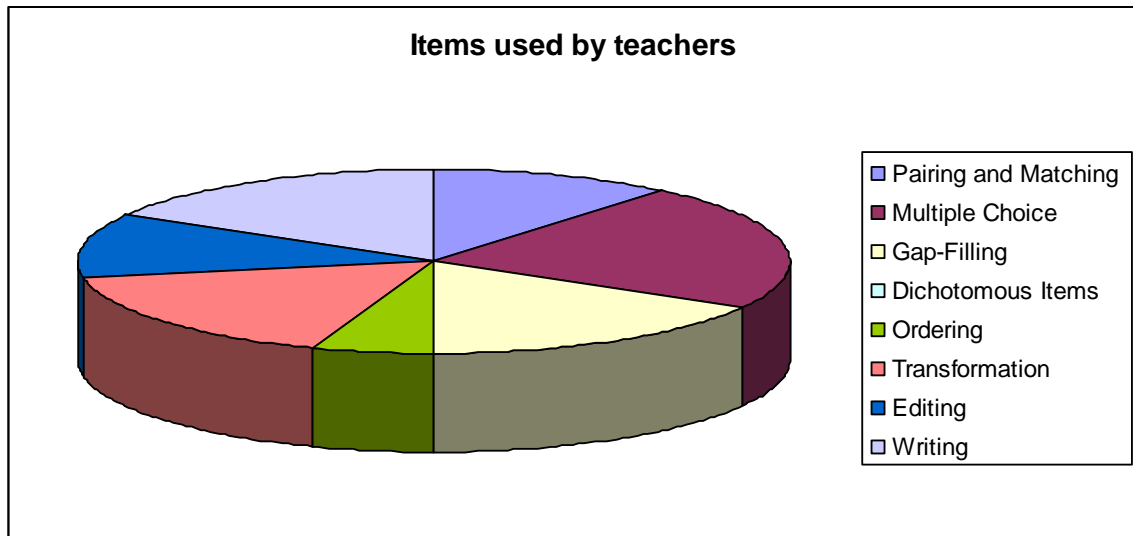
1- Which of the items below do you include in your tests?

- Pairing and Matching (matching responses from the alternatives given)
- Multiple Choice (selecting a response from among alternatives provided)
- Gap-Filling (restoring the missing word in a given passage)
- Dichotomous Items (True/ False- Yes/ No- questions)
- Ordering (arranging “broken” sentences)
- Transformation (changing the structural pattern of a word/ phrase)
- Editing (identifying the errors which have been introduced in a given text)
- Writing

This is the average taken from their responses:

The highest percentage (21%) was revealed in the use of multiple choice items. Still, three other items, which are gap-filling, ordering and writing, were selected with a 17% each. Dichotomous items are used only 6% of the times and both editing and transformation showed a result of 11%. An average of 25% of the teachers interviewed explained that writing is used only when the students ask

for this type of practice since, according to these teachers, most adults refuse to write.



Question 2

Items used for testing grammar

2- Which of these items do you include when you test your students' knowledge of grammar?

- Pairing and Matching (matching responses from the alternatives given)
- Multiple Choice (selecting a response from among alternatives provided)
- Gap-Filling (restoring the missing word in a given passage)
- Dichotomous Items (True/ False- Yes/ No- questions)
- Ordering (arranging "broken" sentences)

- Transformation (changing the structural pattern of a word/ phrase)
- Editing (identifying the errors which have been introduced in a given text)
- Writing

Once teachers were asked about the exercises they use in general, special attention was given to the items chosen to assess their students' knowledge of grammar.

Most teachers, three out of four, chose multiple choice, gap-filling and writing tasks to check their students' grammar performance. Two teachers also include exercises where students have to change the pattern of a word and only one teacher includes matching and editing activities. Neither dichotomous items nor ordering tasks are used by these teachers to test their students. Surprisingly, their choices were in keeping with the types of items constructed in the grammar test.

Question 3

Items which best reflect students' knowledge of grammar

Gap filling and open tasks are considered to be the best activities to test students' knowledge of grammar. Two teachers also agree that transformation items which require students to change the pattern of words and phrases also

reflect students' acquired grammar. Only one teacher believes that editing and multiple choice exercises might reveal students' real knowledge of the grammar studied in class.

The data collected in this question was analyzed below considering each of the answers given by the teachers interviewed.

Teacher I

This teacher explained during the interview that she works with different types of tasks in class and tests her students with a variety of items too. She said that she feels that some activities offer more difficulties than others. One of the examples given was the use of gap filling exercises because, in her opinion, they require not only grammar knowledge, but also good comprehension on the part of the student as well as other structures and strategies to fill in the blanks successfully.

Writing assignments, which are frequently used by this teacher, are meant to be useful as a follow up of the students' learning process. This type of tasks help her find out not only what learners have acquired but also what they are able to produce on their own and what still needs to be practised.

According to this teacher's experience, gap filling, editing and transformation items bring more difficulties to her students because they require prior training and a thorough explanation of what is expected in each exercise. In her experience, the more students practise, the better they will perform.

Teacher II

This teacher believes that gap filling, transformation, editing and writing are the best task to test grammar. Her choices usually depend on the group. As she teaches in two of the groups under study, she was asked if there were any differences between them and she pointed out that in group C, students are more open to writing tasks which helps her identify the type of mistakes that affect the transmission of the message.

Gap filling offers greater difficulties since, for this teacher, students need to understand the text, the overall idea and as it is not their own production, they are forced to use grammar or lexical items that they would not use otherwise. This is why, this teacher explains that gap filling and writing items reflect the students' knowledge of grammar much better than other items. She also believes that both open and closed items require previous training. This became clearer to her when her students were first assessed with the TOEIC exam and the scores they got were very different from the results shown the

following week when they were tested again but by her with familiar activities that are used in class.

Teacher III

Her grammar tests usually include gap filling, transformation and writing items because in this way she can assess grammar and students' written understanding. However, transformation exercises, in her opinion, are harder because learners must integrate all their skills in order to complete them correctly.

In this teacher's experience, students are unlikely to obtain good results in most items without previous training except for activities where alternatives are given and guessing the right answer might help the test taker get the expected outcome.

Teacher IV

This teacher uses all the test items listed except for writing. She explains that ever since she started working with adults, most of them are reluctant to writing tasks and this is the reason why she only includes mechanical exercises to fix the grammar items seen in class and to test her students as well.

She chooses all the items listed but her choice is always limited to the group since all her students study English with a different goal and she provides the tools in the hope of helping them gain autonomy.

Although the challenge in each exercise depends on the level, this teacher is convinced that writing is the most difficult task because it requires knowledge of grammar, vocabulary, register etc.

Students need previous training on all the items, even on writing tasks. Otherwise it would be like a “guessing game” instead of the reflection of what they are able to produce. Adult students go further in their learning process if they are told clearly what they should do since they have different needs, limitations and experiences. Part of this learning process depends on how much practice they get, the more familiar they are with the type of item, the more confidence they gain and the better results they attain when they are tested.

Triangulation of research results

As the purpose of this study is to find out to what extent the type of test item chosen to evaluate grammar influences students' test results an inclusive analysis of each item will be carried out. The triangulation of the results will be explored in the light of the seventeen participants' results and the influence of the groups' perceptions as well as their teachers' experiences.

Group A

Most of the mistakes identified in item one in this group were in the use of verb tenses. In fact, the few mistakes made by student 1 and 3 were identified in this structure and student 3 failed to select correctly all the gaps where verb tenses were being tested. The three participants agree that the alternatives given in this type of items are mostly tricky and difficult. Still, when they were asked to rank different test items, multiple choice exercises were among the easiest activities. Item 4 also tested different structures in a multiple choice format and the choice of verb tenses was also the main difficulty in this group. In two cases, the present continuous was used instead of the present perfect and in one case the participant failed to use the past simple correctly.

Two out of the four mistakes in the second section of the test were in the failure to produce the relative pronoun *who*, which was properly used in items 1 and 2. Unlike the other close items, verb tenses were correct except for one error. This task was ranked as more difficult than multiple choice by all the members of this group and although they are not usually tested with this format, their teacher uses gap filling exercises in class because, as she expressed in the interview, they require not only grammar knowledge, but also a good comprehension.

Although there were some structures which were omitted in the third task, verb tenses were properly used by two test takers who explain in the questionnaire that they often use the language for business trips. The third participant left this part of the test incomplete and although he has been studying English for the same number of years as his/her classmates, he/she only practises English in class which might be a reason why integrating his/her knowledge seems more challenging. The teacher of this group makes her students do writing assignments as a follow up of the students' learning process even if they are not tested in this way later on.

Multiple choice items either in isolated or contextual sentences seemed to have shown the group's weakness in verb tenses. However, surprisingly enough, this structure was correct in most cases where students were asked to produce it rather than identify it as in the case of the third and fourth item. On the other

hand, a recurrent mistake in the second item was not repeated in the other activities. It seems, therefore, that the way the students of this group performed in one task differed from the way they did in another.

Group B

The results in this item were very different since one test taker succeeded in completing most of this part correctly, another member of this group failed to choose the right choice in most of the sentences and the third test taker filled 50% successfully. In fact, their perceptions towards this type of exercise also differ since although they all agree that the alternatives are usually tricky, two participants find this item easy and one of them does not. The teacher in charge of this group firmly believes that training is essential for completing items successfully and it would be interesting to point out that the student who got mainly wrong answers in this item has studied English for only two years since most of his knowledge was acquired travelling.

In spite of the differences in the first exercise, the results obtained in the second item were mainly in the use of “have to” and the relative pronoun “who”. The teacher of this group believes that gap filling offers greater difficulties since students are forced to use grammar or lexical items that they would not use otherwise. However, there were fewer mistakes in this section than in the multiple choice section.

Likewise, the results in this task were also quite similar since all the participants managed to pass on a clear message even though the structures used were very simple and most of the expected grammar items were omitted.

It is worth noticing that this group is usually tested with this type of format and still, the results show differences among the group since one test takers made the same mistakes in this item as in item 1 and the other students completed all the blanks correctly.

Even if the results obtained were uneven in terms of the group, the outcome in the first and last items were similar for each participant and the performance in the second and third item were similar in terms of the level of the group and different if compared to the rest of the formats.

The main difference in these students' experience which might influence their uneven results is that one of the participants has never studied English with continuity; the other has studied regularly and needs the language for work and the last participant studies for future opportunities but does not use the language at work.

Group C

Both participants chose all the alternatives correctly except for the identification of uncountable nouns. In the multiple choice item presented in isolated sentences the same number of mistakes were shown but these mistakes were identified in sentences that tested verb tenses. This type of items was ranked differently since one test taker considers it one of the easiest formats and the other student regards it as one of the most difficult ones.

Gap filling, for these students' teacher, offers greater difficulties since students need to understand the text and the overall idea. However, one of the students ranked it as the easiest format and both students managed to complete all the gaps correctly except for one mistake made by one of the students who failed to complete the gap that tested uncountable nouns.

One of the students succeeded in producing a clear and accurate paragraph using the expected grammar points and the other made many grammar mistakes but none of these errors prevented the participant from transmitting on a clear message. Both participants believe that this is the most challenging format used for testing and they also agree that when a type of item is practised in class, one feels more confident to complete the exercise. Their outcome was positive because as was pointed out by their teacher, these students are more open to writing tasks than group B.

Group D

The most recurrent mistake in this item was in the choice of grammatical categories. This mistake was not repeated in the last item since only one student made two mistakes in the last section and neither of them was in grammatical categories. Only one student in this group believes that the alternatives given in multiple choice exercises are easy. However, most participants managed to complete these two items successfully. In fact, better results were shown in the last item which tested students' knowledge of grammar in sentences in isolation. These students are tested in this way by an external institution but their teacher does not include this type of item in her tests because, in her opinion, it is likely to guess the right answer in activities where alternatives are given.

These participants' learning experience and perceptions towards test items are diverse and although they usually get training in gap filling items as well as in writing tasks, the results obtained in the test were very varied: two participants made a mistake in grammatical categories as in item one, another student failed to complete most of the blanks and the fourth test taker completed all the gaps well. Regarding the writing item, most participants expressed themselves accurately using one to three of the expected grammatical structures and only one student failed to express his/her ideas clearly.

Group E

The student who completed all the tasks successfully did not do the writing section and the test taker who made some mistakes, which were repeated throughout, managed to write a clear and well organized paragraph in spite of some minor mistakes.

One of the student's grammar weaknesses were clearly shown in this test since the same mistakes were repeated throughout in spite of the different contexts and formats which distinguish one exercise from the other. Verb tenses, grammatical categories and the identification of uncountable nouns presented difficulties in both item one and two.

Writing, for both students, seems to be the hardest activity and they agree with their teacher in that previous practice is essential in the outcome of the tests. One of these test takers managed to produce a clear and well organized paragraph along with some expected grammar structures. However, most verb tenses were used incorrectly. The other participant in this group did not complete this task but filled all the gaps in the test perfectly well.

Group F

The types of mistakes identified in the first item were varied and surprisingly this format was ranked as one of the most complex by two participants and the alternatives given were described as tricky by all the students. The teacher of this course uses a variety of items to test her students, especially multiple choice because it is a good way to fix the grammar items seen in class and because students are tested in this way by the external teachers.

Although gap filling seems challenging for this group, all the gaps were properly completed by two test takers and the other student only made one mistake in the choice of relative pronouns, which was not a mistake made in the first item by any of these students.

The last item showed different results from the first exercise. Two out of the three students in this group completed all the gaps with the correct word and the third participant made two errors which were different from the ones made in previous items.

This teacher uses all the test items listed except for writing. She explains that ever since she started working with adults, most of them are reluctant to writing tasks and this is the reason why she only includes mechanical exercises. Yet, the three participants succeeded in producing a clear and accurate paragraph

using some of the expected grammar points. Students as well as their teacher believe that writing is the most difficult task and that previous training on all the items, even on writing tasks, is needed to do well.

Test Items

The data collected were explored below in the light of the results obtained in each of the test items.

Item 1

The results obtained in this first item, which was designed as a multiple choice exercise presented in a short paragraph including eight gaps to be completed with one out of four alternatives, revealed that although there were seven different structures assessed, the participants in this study managed to complete an average of 70% correctly.

The lowest score was in the gap that tested grammatical categories. The options presented included one correct alternative and one possible choice. The latter was not appropriate due to the context in which it appeared and most test takers seemed to have been confused and chose this alternative. The highest percentage appeared in relative clauses since there were 88% of correct answers.

The eight gaps in this multiple choice item were completed by 100% of the students who agree that most options given in this type of activities are tricky. Four people who answered the questionnaire believe that they are so because some confusing or unknown words appear and make the choice harder to make. When students were asked to rank the different activities, 40% ranked multiple choice items as easier than writing a composition. Thirteen students also believe that the type of previous training they get before the test influences the results they obtain.

Item 2

In order to fulfill this task successfully, participants needed both comprehension of the paragraph and enough knowledge of grammar to produce the expected response. A lower percentage of blanks was completed successfully if compared to item one and 4% of the gaps were not filled. These results seem to give evidence to the fact that when test takers are not given options, finding the right word might be more difficult than picking it from a given set of alternatives. However, only five participants out of seventeen ranked this activity as one of the most difficult whereas teachers consider it one of the most challenging along with writing. If compared to the first item and bearing in mind the absence of choices to fill in the gaps, the number of correct answers wasn't so low since 59% of the answers were correct.

What is worth noticing is that most of the mistakes made in this item were in the use of relative pronouns and the recognition of uncountable nouns. The former was a structure which was used 88% of the times correctly in the first item but most students failed to produce it in the gap it was required. Also, many test takers completed one of the gaps with the word “many” which shows that they were aware of the type of word that was necessary but failed to realize that “much” was more suitable due to the presence of the noun traffic. It seems, therefore, that the most recurrent mistakes in this activity do not match with the errors made in the first item since the gaps that tested tenses, categories and modal verbs were properly filled by most test takers in this item.

Item 3

In this open task, 45% of the implicitly expected structures were properly used whereas 51% of these structures were omitted or in most cases replaced by simpler sentences.

The use of verb tenses and modal verbs was correct in over 50% of the cases and 80% of the students who performed well in the use of these structures are people who use the language frequently for business purposes. Relative clauses, on the other hand, which were properly chosen in the first item but

presented many difficulties in the second, were mainly omitted in this task by most test takers.

Two out of the seventeen learners left this section of the exam incomplete even though their teachers frequently give them this type of assignments. Surprisingly, however, both got the highest results in the rest of the items. Only one student failed to pass on a clear message, ten students expressed themselves clearly even though they resorted to simpler structures and four test takers succeeded in completing the task and also used most of the expected grammar structures.

Within the group of students who produced a piece of writing which included the expected grammar structures, only 4% of mistakes were identified and surprisingly enough, one of the few test takers who managed to integrate the expected language in this response, attended the course where the teacher prefers not to give them writing tasks.

Writing, for most of the participants in this study, is the hardest activity and they all agree with their teacher in that previous practice is essential in the outcome of the tests except for three students who think that writing skills depend on the person's personal skills rather than on practice.

Item 4

This item was included especially because it is the format used by the external institutions that assess the participants of this study. The results obtained were in keeping with the expected outcome since there was only a 15% of wrong answer and none of the blanks were left unanswered. In most of the groups, structures that were not identified in the first item, omitted in the second item and avoided in the third item were well chosen in this part. Even if the results are compared to the first item, which was also a multiple choice but in context, the results obtained were higher in this section.

The four teachers interviewed said that their students are familiar with multiple choice exercises because four of these groups are tested in this format in the TOEIC exam and two groups are tested in this way by an external teacher. However, three of these teachers, who do include multiple choice items in their grammar assessments, do not regard this item as one which best reflects students' knowledge.

CHAPTER 5

CONCLUSIONS

This study intended, as stated in the research question, to find out to what extent the results obtained by in-company intermediate students are influenced by the type of written item chosen by the teacher in the construction of grammar tests.

Based on the analysis of the data collected through a written grammar test, questionnaires for students and interviews to teachers, the findings of this investigation seem to indicate that the types of items used to evaluate students knowledge of grammar influenced the results that the participants in this investigation obtained.

As far as the first hypothesis is concerned, which stated that close-item exercises seem easier than open-item tasks, it can be said that the outcome of the tests as well as the teachers and the students' perceptions showed that this is so. Moreover, when asked to signal from a list the most difficult test items, 90% of the participants managed to rank open tasks as the hardest whereas multiple choice, matching and ordering or dichotomous items were always ranked among the least challenging. Teachers showed an accurate perception of the degree of difficulty their students had in each of these items. However, there were two

cases where the outcome of the students' writing performance was better than their choices in close-items. All in all, the results attained in closed-items were over a 70% of correct answers whereas an average of 50% was obtained in the open tasks. This being the results of the study, the first hypothesis might be said to be right.

With respect to the second hypothesis, which attempted to affirm that close-item exercises in a given context may appear clearer for the students to complete than exercises in isolation, the assumption seems to be wrong since 95% of the students made fewer mistakes in the item where sentences were presented in isolation than in the multiple-choice sentences in a given context. On the other hand, when asked what type of items students were used to doing, all the teachers stated that activities including isolated sentences were used to test grammar because participants are assessed in this way even though one of the teachers interviewed openly expressed that this format does not truly reflect the students' knowledge of grammar. Most teachers and students agree that previous training might influence their performance in a given test item. Therefore, if the students in this investigation get further practice on isolated sentences than they do in contextual sentences, the results might have been influenced also by their experience with this test format. This seems, to a large extent, to prove the second hypothesis not conclusively wrong.

The ambiguity that resulted from the conclusion of the second hypothesis seemed to pave the way for the third hypothesis, which stated that closed-item tasks seemed to require previous practice rather than knowledge. There seemed to be some incongruity between what teachers and students think and what students were actually able to do. All the teachers admitted that practice is essential for students to perform well in an exam and in fact, most students performed well in the last item, which is the format often used to test them. Yet 50% of the students who got outstanding results in the writing task never do any assignment of this type in class and two of the students who obtained top marks in close-items did not complete the writing activity even though their teachers stated that they frequently practice writing in class. Moreover, one teacher and three students pointed out that the learners' performance depends on personal skills rather than on practice. As the results of the test showed, most of the participants who managed to write an accurate paragraph, regardless of the structures used, were those whose background was in keeping with a person who uses the language regularly for work. 35% of these learners used structures in the writing task that were not properly used in close-item selection whereas the remaining 60% recognized the structures but were unable to integrate them in the writing item. Test takers, therefore, seem to benefit from further practice. This being the case, the third hypothesis seems to have been proved right.

The fourth hypothesis expressed that open-item tasks aimed at checking knowledge of grammar may not always reflect the use of intended grammar

structures. The outcome of the writing task showed that over 50% of the implicitly expected structures were omitted or replaced by simpler language whereas 96% of the blanks in the gap filling exercise, which forces the test taker to produce the expected word, were completed even though 40% were unsuitable words. In 30% of the cases, structures that were well used in these types of items were chosen incorrectly in close-item exercises. It seems that, at this level, students can handle basic structures to pass on the intended message and new or more complex structures are avoided or omitted. Yet, when used, very few mistakes were made.

If open test items are considered what Oller termed (1989) pragmatic items, such as gap-filling reading items, where students focus on both knowledge of the formal linguistic features and on the way language is used to achieve communication, the second hypothesis would be wrong. Still, if, as Frenco (2000) states, open items are regarded as activities that do not require a specific answer, the data collected rendered seemed to prove this hypothesis right.

All in all, it can be concluded that the type of test item chosen to evaluate in-company intermediate students' knowledge of grammar in the written medium appeared to have influenced their test results. As Bachman (2004) explained, tests are used to make inferences about individuals' language ability and to make decisions about students. Yet, the population studied in this investigation, seemed to spend long hours practising closed items in the hope of exceeding the

scores taken at the beginning of the year even though most teachers admitted that this type of activities do not reflect students' interactive and productive skills but only the recognition of what is grammatically correct or incorrect. It can be concluded, therefore, that when students were tested on grammar through close-item selection, most test takers were able to take a passing mark whereas the same participants failed to get over 70% of correct answers when tested on the same structures but through an open item task. As language tests seem to be an important part of education, it is important to be aware of the importance of test construction in order to obtain reliable information from them because as Frendo clearly states: "Evaluation is of critical importance in supporting and shaping how teaching is done (2009:123)".

5.1. Limitations and suggestions for future research

The small number of students and teachers in the investigation limit the possibilities of generalizing the findings from this study. Students were chosen according to the results they got on their placement test, which was mainly designed with a closed-item test and not all the types of items were included in the test used as content analysis. This would bring to focus another variable in the investigation. In addition, it is not clear whether the conclusions of the study could apply to groups of students who study in company but who are not tested by external institutions. For these reasons as well, this investigation is

exploratory and does not aim at establishing any absolute conclusions on the issue under study.

Future lines of research could be conducted on a larger scale, as this study analyzed only six groups of students and their teachers of English and some other results and conclusions might arise from a large-scale investigation. Also, the investigation could be carried out with students of a different language level or in a different context such as school or university. In addition, teachers' training in the construction of language tests as well as other factors at stake in the design, the testing and the scoring process could still be studied.

REFERENCES

Alderson, J.C. & Beretta, A. (1992) *Evaluating second language education*. Cambridge; Cambridge University Press.

Alderson, J.C, Clapham, C. & Wall, D. (2005). *Language test construction and evaluation*. Cambridge; Cambridge University Press

Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.

Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bachman, L. F. (1996). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Brown, H.D. (2000). *Teaching by principles*. New York: Longman.

Donna, S. (2000). *Teach business English*. Cambridge: Cambridge University Press.

Ellis, M and Johnson, C. (2000). *Teaching business English*. Oxford: Oxford University Press.

Evans T.D and St. John M. *Developments in ESP*. Cambridge: Cambridge University Press.

Frendo, E. (2009). *How to teach business English*. New York: Longman.

Heaton, J.B. (1988). *Writing English language tests*. Harlow, Essex: Longman.

Hewings M. and Nickerson C. (2000). *Business English: Research into practice*. London: Longman.

Hughes, A. (2003) *Testing for language teachers*. Cambridge: Cambridge University Press.

Groundlud, N.E., (1971) *Measurement and evaluation in teaching*. New York: Macmillan.

McNamara, T (2007). *Language testing*. Oxford: Oxford University Press.

McNamara, T. and Deane, D. (1995). *Self-assessment activities: Toward autonomy in language learning*. TESOL Journal.

Oller, J. (1989). *Language tests at school*. London: Longman.

O'Malley, J.M. and Valdez Pierce, L. (1995). *Authentic assessment for English language learners. Practical Approaches for Teachers*. Addison- Wesley Publishing Company.

Purpura, J. E. (2004). *Assessing grammar*. Cambridge: Cambridge University Press.

Weigle S.C (2002). *Assessing writing*. Cambridge: Cambridge University Press.

APPENDIX

APPENDIX I

Research Instruments

Grammar Test

1- Read the text below and decide which answer best fits each space: **A, B, C** or **D**.

Google

Many companies 1-_____ to make money from the Internet in these last years but it seems that people did not buy enough to keep the dot.com retailers in business. Advertisers 2-_____ pay to get their products at the top of the list when people search for information. Sixty thousand advertisers, for example, use Google's services 3-_____ worldwide. This is the world's 4-_____ search engine 5-_____ has AOL and EarthLink as its clients. If Google is 5- interested in staying in first place, it will have to keep prices low. The Managing Director explained that advertisers give the company 6-_____ profit so they don't make enough money from them. But as there is strong competition these days, the company 7-_____ to offer a good service. Therefore, Google is trying to provide its customers with services 8-_____ are different from their competitors.

- 1- A- try B- has tried C- have tried D- is trying
- 2- A- ought to B- should C- have to D- has to
- 3- A- effective B- more effective C- effectively D- more effectively
- 4- A-bigger B- big C- biggest D- the biggest
- 5- A- where B- who C- where D- which
- 6- A- much B- few C- many D- little
- 7- A- work B- had worked C- worked D- is working
- 8- A- who B- whose C- which D- where

2- Read the paragraph below and fill in the blanks with a suitable word.

Andrew Miwa, China's Porsche salesman, is ready for the _____ important boom in the car business since 1920. In Beijing alone, about 4,000 new cars arrive every week and 600,000 new drivers _____ get licenses next year. However, Mr. Miwa, _____ goes to work by car every day, has realized that the number of cars _____ causing great problems. The government has _____ destroy many areas to change roads into highways to avoid so _____ traffic. "People love _____ to work by car because it is _____ but something should be done" Miwa says.

3- Your company has decided that the use of the Internet should be prohibited during working hours. Write an e-mail to a friend telling him why the company has come to this decision and describe how the situation has changed.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

4- Read the following sentences and choose the right option.

1- Julia English for ten years.

- a- have studied b- studies c- is studying d- has studied

2- Steve Jobs, is the owner of Apple, is about to retire.

- a- whom b- who c- that d- whose

3- Katherine arrived early because she had too..... work.

- a- few b- little c- much d-many

4- I've got three brothers. The lives in Europe.

- a- elder b- older c- elderly d- eldest

5- Tim Burton is my uncle. He is a lawyer.

- a- success b- successful c- successfully d- more successful

6- There is a lot of food for dinner so you bring anything.

- a- shouldn't b- mustn't c- can't d- don't have to

7- Paul in the Finance Department since he graduated.

- a- has worked b- had worked c- have worked d- works

8- My boss a project at the moment to set up a new business abroad.

- a- makes b- is making c- made d- has made

APPENDIX II

Students' questionnaire

1- Las alternativas que se dan en un examen de respuestas múltiples suelen ser:

- capciosas
- complejas
- sencillas
- otro

2- ¿Cuáles de los siguientes ejercicios te resulta más sencillo? Numerar del 1 al 7 (siete es el más complejo)

- Unir con flechas palabras, frases o respuestas.
- Llenar los espacios vacíos de un texto con una palabra o frase.
- Elegir una respuesta correcta entre varias alternativas.
- Responder preguntas que pueden ser verdaderas o falsas.
- Ordenar palabras para formar oraciones.
- Transformar palabras para completar correctamente un contexto dado.
- Identificar y corregir errores gramaticales en un texto.
- Escribir una composición.

3- En la evaluación de gramática ¿Cómo te resultaron los ejercicios de respuestas múltiples?

- más fáciles que escribir una composición.
- más difíciles que escribir una composición.
- igual de complejos como escribir una composición.

4- ¿Cuándo te sentís más seguro en la elección de la respuesta?

- cuando ejercitan los temas en un determinado formato en clase.
- hay formatos que te resultan naturalmente mas sencillos.
- depende del tema, varía la dificultad del formato

APPENDIX III

Students' background

a- ¿Cuántos años estudio ingles?

1 a 2 años 3 a 4 años 5 a 6 años mas de 6 años

b- ¿Realizó estos estudios con continuidad?

si no

c-¿Qué tipo de cursos ha tomado?

particular grupales ambos otro

d- ¿Dónde ha estudiado?

en el colegio en un instituto en el trabajo otro

e- ¿Qué tipo de exámenes ha rendido?

de nivel de fin de curso internacionales otro

f- ¿Con cuánta gente ha compartido las clases de inglés?

menos de 5 entre 5 y 10 mas de 10 mas de 20

g- ¿En qué contextos utiliza ud el inglés?

solo en clase en el trabajo viajes de negocios otro

APPENDIX IV

Teachers' interview

1- Which of the items listed below do you use in class?

- Pairing and Matching (matching responses from the alternatives given)
- Multiple Choice (selecting a response from among alternatives provided)
- Gap-Filling (restoring the missing word in a given passage)
- Dichotomous Items (True/ False- Yes/ No- questions)
- Ordering (arranging “broken” sentences)
- Transformation (changing the structural pattern of a word/ phrase)
- Editing (identifying the errors which have been introduced in a given text)
- Writing

2-Which of these items do you include when you test your students' knowledge of grammar?

- Pairing and Matching (matching responses from the alternatives given)
- Multiple Choice (selecting a response from among alternatives provided)
- Gap-Filling (restoring the missing word in a given passage)
- Dichotomous Items (True/ False- Yes/ No- questions)

- Ordering (arranging “broken” sentences)
- Transformation (changing the structural pattern of a word/ phrase)
- Editing (identifying the errors which have been introduced in a given text)
- Writing

3 – Why do you prefer these items? (They are easier to grade, they are objective, they also test comprehension, other learning strategies are tested etc)

4- In your experience, which of these items offer greater difficulties? Why?

5- Which, in your opinion, reflect better your students’ knowledge of grammar?

- Pairing and Matching (matching responses from the alternatives given)
- Multiple Choice (selecting a response from among alternatives provided)
- Gap-Filling (restoring the missing word in a given passage)
- Dichotomous Items (True/ False- Yes/ No- questions)
- Ordering (arranging “broken” sentences)
- Transformation (changing the structural pattern of a word/ phrase)
- Editing (identifying the errors which have been introduced in a given text)
- Writing

6- Do you think the items mentioned above require previous training? Why?

APPENDIX V

Students' questionnaire

Questions 1

Student	A	B	C	D
1		X		
2		X		
3	X			
4	X			
5	X			confusing
6	X			
7		X		
8	X			
9			X	
10	X			
11	X			
12	X			unfamiliar
13	X			
14	X			
15	X			unknown
16	X			
17	X			confusing

Questions 2

Student	A	B	C	D	E	F	G	H
1	1	3	4	6	2	5	0	7
2	1	3	4	6	0	7	2	5
3	3	7	5		4		5	7
4	1	6						7
5	1	3		1	5	2	4	3
6							7	6
7	1	5					6	7
8		2	1			5	6	7
9	1	4	2			6	5	7
10	4	3	5	2	1	6		7
11	1	2	4	3		7	6	5
12	1	5	4	2	3		6	7
13	1	6	5				7	2
14		5	4				6	7
15	1	6	3	2	4	7		5
16	1	3	2			5	6	7
17	4	7	1			5	6	7

Questions 4

Student	A	B	C
1	X		
2		X	
3	X		
4	X		
5			X
6		X	
7	X		
8	X		
9	X		
10	X		
11	X		
12		X	
13	X		
14	X		
15	X		
16	X		
17	X		

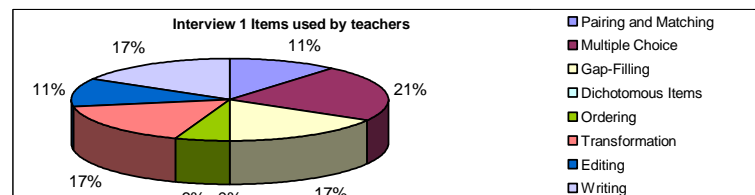
Questions 3

Student	A	B	C
1	X		
2	X		
3	X		
4			X
5	X		
6		X	
7	X		
8	X		
9	X		
10			X
11	X		
12	X		
13		X	
14	X		
15	X		
16	X		
17	X		

Teachers' interview

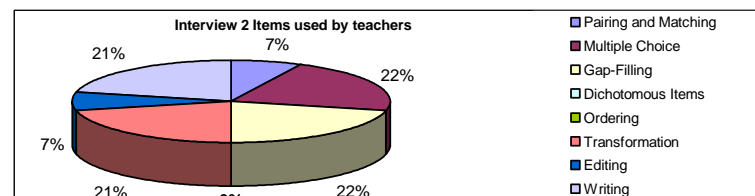
Interview 1

Type of Items	Teacher A	Teacher B	Teacher C	Teacher D	Total
Pairing and Matching	1			1	2
Multiple Choice	1	1	1	1	4
Gap-Filling	1	1		1	3
Dichotomous Items					0
Ordering	1				1
Transformation	1	1		1	3
Editing			1	1	2
Writing		1	1	1	3



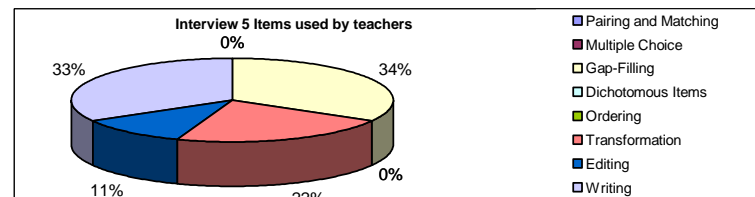
Interview 2

Type of Items	Teacher A	Teacher B	Teacher C	Teacher D	Total
Pairing and Matching	1				1
Multiple Choice	1		1	1	3
Gap-Filling	1	1		1	3
Dichotomous Items					0
Ordering					0
Transformation	1	1		1	3
Editing				1	1
Writing		1	1	1	3



Interview 5

Type of Items	Teacher A	Teacher B	Teacher C	Teacher D	Total
Pairing and Matching					0
Multiple Choice					0
Gap-Filling	1	1		1	3
Dichotomous Items					0
Ordering					0
Transformation	1			1	2
Editing				1	1
Writing	1	1		1	3



Data Matrix: item results per structure

Structure	Results	Absolute Value				Percentages				Max	Min	Avg
		Item 1	Item 2	Item 3	Item 4	Item 1	Item 2	Item 3	Item 4			
Tenses	Correct	13,5	10	13	15	79%	59%	76%	88%	88%	59%	76%
	Incomplete	0	0	2	0	0%	0%	12%	0%	12%	0%	3%
	Incorrect	3,5	7	4	2	21%	41%	24%	12%	41%	12%	24%
	Complete	17	17	15	17	100%	100%	88%	100%	100%	88%	97%
Modals	Correct	13	9	12	14	76%	53%	71%	82%	82%	53%	71%
	Incomplete	0	1	5	0	0%	6%	29%	0%	29%	0%	9%
	Incorrect	4	8	5	3	24%	47%	29%	18%	47%	18%	29%
	Complete	17	16	12	17	100%	94%	71%	100%	100%	71%	91%
Grammatical Categories	Correct	10	13	4	15	59%	76%	24%	88%	88%	24%	62%
	Incomplete	0	1	12	0	0%	6%	71%	0%	71%	0%	19%
	Incorrect	7	4	13	2	41%	24%	76%	12%	76%	12%	38%
	Complete	17	16	5	17	100%	94%	29%	100%	100%	29%	81%
Comparatives & Superlatives	Correct	13	12	7	15	76%	71%	41%	88%	88%	41%	69%
	Incomplete	0	1	10	0	0%	6%	59%	0%	59%	0%	16%
	Incorrect	4	5	10	2	24%	29%	59%	12%	59%	12%	31%
	Complete	17	16	7	17	100%	94%	41%	100%	100%	41%	84%
Relative Clauses	Correct	15	10	2	14	88%	59%	12%	82%	88%	12%	60%
	Incomplete	0	0	15	0	0%	0%	88%	0%	88%	0%	22%
	Incorrect	2	7	15	3	12%	41%	88%	18%	88%	12%	40%
	Complete	17	17	2	17	100%	100%	12%	100%	100%	12%	78%
Countable & Uncountable	Correct	11	6	8	14	65%	35%	47%	82%	82%	35%	57%
	Incomplete	0	1	8	0	0%	6%	47%	0%	47%	0%	13%
	Incorrect	6	11	9	3	35%	65%	53%	18%	65%	18%	43%
	Complete	17	16	9	17	100%	94%	53%	100%	100%	53%	87%

Data Matrix: Test results per student, group and teacher.

Group	Teacher	Student	Items	Tenses	Tenses Completed (Y/N)	Modals	Modals Completed (Y/N)	Grammatical Categories	Grammatical Completed (Y/N)	Comparatives & Superlatives	C&S Completed (Y/N)	Relative Clauses	Relative Completed (Y/N)	Countable & Uncountable	Count & uncount Completed (Y/N)	Total Correct	Total Incorrect	Answered Total	Total Unanswered	Answer Percentage	Unanswered Percentage	Percentage Correct	Percentage Incorrect
A	I	1	Gap filling in context	1	1	0	1	1	1	1	1	1	1	0	1	4,0	2,0	6,0	-	100%	0%	67%	33%
A	I	2	Gap filling in context	0	1	1	1	1	1	1	1	1	1	0	1	4,0	2,0	6,0	-	100%	0%	67%	33%
A	I	3	Gap filling in context	1	1	1	1	1	1	1	1	0	1	0	1	4,0	2,0	6,0	-	100%	0%	67%	33%
B	II	4	Gap filling in context	0	1	0	1	0	1	0	1	1	1	0	1	1,0	5,0	6,0	-	100%	0%	17%	83%
B	II	5	Gap filling in context	0,5	1	0	1	1	1	1	1	0	1	1	1	3,5	2,5	6,0	-	100%	0%	58%	42%
B	II	6	Gap filling in context	0,5	1	0	1	0	1	1	1	1	1	1	1	3,5	2,5	6,0	-	100%	0%	58%	42%
C	II	7	Gap filling in context	0,5	1	1	1	1	1	1	1	0	1	1	1	4,5	1,5	6,0	-	100%	0%	75%	25%
C	II	8	Gap filling in context	1	1	1	1	1	1	1	1	1	1	1	1	6,0	-	6,0	-	100%	0%	100%	0%
D	III	9	Gap filling in context	0,5	1	0	1		0	0	1	1	1	0	1	1,5	3,5	5,0	1,0	83%	17%	30%	70%
D	III	10	Gap filling in context	0	1	1	1	1	1	1	1	1	1	1	1	5,0	1,0	6,0	-	100%	0%	83%	17%
D	III	11	Gap filling in context	1	1	1	1	1	1	1	1	1	1	0	1	5,0	1,0	6,0	-	100%	0%	83%	17%
D	III	12	Gap filling in context	0	1	0	1	1	1	0	1	0	1	0	1	1,0	5,0	6,0	-	100%	0%	17%	83%
E	III	13	Gap filling in context	0,5	1		0	1	1		0	0	1		0	1,5	1,5	3,0	3,0	50%	50%	50%	50%
E	III	14	Gap filling in context	1	1	1	1	1	1	1	1	0	1	1	1	5,0	1,0	6,0	-	100%	0%	83%	17%
F	IV	15	Gap filling in context	1	1	1	1	0	1	0	1	0	1	0	1	2,0	4,0	6,0	-	100%	0%	33%	67%
F	IV	16	Gap filling in context	0,5	1	1	1	1	1	1	1	1	1	0	1	4,5	1,5	6,0	-	100%	0%	75%	25%
F	IV	17	Gap filling in context	1	1	0	1	1	1	1	1	1	1	0	1	4,0	2,0	6,0	-	100%	0%	67%	33%
			Total Gap Filling	10	17	9	16	13	16	12	16	10	17	6	16	60	38	98	4	96%	4%	61%	39%
A	I	1	Multiple choice (isolated)	1	1	1	1	1	1	1	1	1	1	1	1	6,0	-	6,0	-	100%	0%	100%	0%
A	I	2	Multiple choice (isolated)	0,5	1	1	1	1	1	1	1	1	1	0	1	4,5	1,5	6,0	-	100%	0%	75%	25%
A	I	3	Multiple choice (isolated)	1	1	1	1	1	1	1	1	1	1	0	1	5,0	1,0	6,0	-	100%	0%	83%	17%
B	II	4	Multiple choice (isolated)	0,5	1	1	1	1	1	1	1	1	1	1	1	5,5	0,5	6,0	-	100%	0%	92%	8%
B	II	5	Multiple choice (isolated)	1	1	1	1	1	1	1	1	1	1	1	1	6,0	-	6,0	-	100%	0%	100%	0%
B	II	6	Multiple choice (isolated)	1	1	0	1	1	1	1	1	1	1	0	1	4,0	2,0	6,0	-	100%	0%	67%	33%
C	II	7	Multiple choice (isolated)	1	1	0	1	0	1	0	1	1	1	1	1	3,0	3,0	6,0	-	100%	0%	50%	50%
C	II	8	Multiple choice (isolated)	1	1	1	1	1	1	1	1	1	1	1	1	6,0	-	6,0	-	100%	0%	100%	0%
D	III	9	Multiple choice (isolated)	1	1	1	1	1	1	1	1	1	1	1	1	6,0	-	6,0	-	100%	0%	100%	0%
D	III	10	Multiple choice (isolated)	1	1	1	1	1	1	1	1	1	1	1	1	6,0	-	6,0	-	100%	0%	100%	0%
D	III	11	Multiple choice (isolated)	1	1	1	1	1	1	1	1	0	1	1	1	5,0	1,0	6,0	-	100%	0%	83%	17%
D	III	12	Multiple choice (isolated)	1	1	1	1	1	1	1	1	0	1	1	1	5,0	1,0	6,0	-	100%	0%	83%	17%
E	III	13	Multiple choice (isolated)	0	1	1	1	1	1	1	0	1	1	1	1	4,0	2,0	6,0	-	100%	0%	67%	33%
E	III	14	Multiple choice (isolated)	1	1	1	1	1	1	1	1	1	1	1	1	6,0	-	6,0	-	100%	0%	100%	0%
F	IV	15	Multiple choice (isolated)	1	1	1	1	0	1	1	1	1	1	1	1	5,0	1,0	6,0	-	100%	0%	83%	17%
F	IV	16	Multiple choice (isolated)	1	1	1	1	1	1	1	1	1	1	1	1	6,0	-	6,0	-	100%	0%	100%	0%
F	IV	17	Multiple choice (isolated)	1	1	0	1	1	1	0	1	1	1	1	1	4,0	2,0	6,0	-	100%	0%	67%	33%
			Multiple choice (isolated)	15	17	14	17	15	17	15	17	14	17	14	17	87	15	102	0	100%	0%	85%	15%

Data Matrix: Test results per student, group and teacher. Cont.

Group	Teacher	Student	Items	Tenses	Tenses Completed (Y/N)	Modals	Modals Completed (Y/N)	Grammatical Categories	Grammatical Completed (Y/N)	Comparatives & Superlatives	C&S Completed (Y/N)	Relative Clauses	Relative Completed (Y/N)	Countable & Uncountable	Count & uncount Completed (Y/N)	Total Correct	Total Incorrect	Answered Total	Total Unanswered	Answer Percentage	Unanswered Percentage	Percentage Correct	Percentage Incorrect
A	I	1	Multiple choice in context	0,5	1	1	1	1	1	1	1	1	1	1	1	5,5	0,5	6,0	-	100%	0%	92%	8%
A	I	2	Multiple choice in context	0	1	1	1	0	1	0	1	1	1	0	1	2,0	4,0	6,0	-	100%	0%	33%	67%
A	I	3	Multiple choice in context	0,5	1	1	1	0	1	1	1	1	1	0	1	3,5	2,5	6,0	-	100%	0%	58%	42%
B	II	4	Multiple choice in context	0,5	1	0	1	0	1	0	1	0,5	1	0	1	1,0	5,0	6,0	-	100%	0%	17%	83%
B	II	5	Multiple choice in context	1	1	1	1	0	1	1	1	1	1	1	1	5,0	1,0	6,0	-	100%	0%	83%	17%
B	II	6	Multiple choice in context	0,5	1	0	1	0	1	1	1	1	1	0	1	2,5	3,5	6,0	-	100%	0%	42%	58%
C	II	7	Multiple choice in context	1	1	1	1	0	1	1	1	1	1	1	1	5,0	1,0	6,0	-	100%	0%	83%	17%
C	II	8	Multiple choice in context	1	1	1	1	1	1	1	1	1	1	1	1	6,0	-	6,0	-	100%	0%	100%	0%
D	III	9	Multiple choice in context	1	1	1	1	1	1	1	1	1	1	1	1	6,0	-	6,0	-	100%	0%	100%	0%
D	III	10	Multiple choice in context	1	1	1	1	1	1	1	1	1	1	1	1	6,0	-	6,0	-	100%	0%	100%	0%
D	III	11	Multiple choice in context	0,5	1	1	1	1	1	1	1	1	1	0	1	4,5	1,5	6,0	-	100%	0%	75%	25%
D	III	12	Multiple choice in context	1	1	0	1	1	1	0	1	1	1	1	1	4,0	2,0	6,0	-	100%	0%	67%	33%
E	III	13	Multiple choice in context	1	1	1	1	1	1	1	1	0	1	0	1	4,0	2,0	6,0	-	100%	0%	67%	33%
E	III	14	Multiple choice in context	1	1	1	1	1	1	1	1	1	1	1	1	6,0	-	6,0	-	100%	0%	100%	0%
F	IV	15	Multiple choice in context	1	1	0	1	0	1	0	1	1	1	1	1	3,0	3,0	6,0	-	100%	0%	50%	50%
F	IV	16	Multiple choice in context	1	1	1	1	1	1	1	1	1	1	1	1	6,0	-	6,0	-	100%	0%	100%	0%
F	IV	17	Multiple choice in context	1	1	1	1	1	1	1	1	0,5	1	1	1	5,5	0,5	6,0	-	100%	0%	92%	8%
			Multiple choice in context	13,5	17	13	17	10	17	13	17	15	17	11	17	75,5	26,5	102	0	100%	0%	74%	26%
A	I	1	Writing	1	1	1	1	1	1	0	0	0	0	1	1	4,0	-	4,0	2,0	67%	33%	100%	0%
A	I	2	Writing	1	1		0		0		0		0		0	1,0	-	1,0	5,0	17%	83%	100%	0%
A	I	3	Writing	1	1	1	1	1	0	1	1		0		0	3,0	-	3,0	3,0	50%	50%	100%	0%
B	II	4	Writing	1	1	1	1		0	1	1		0		0	3,0	-	3,0	3,0	50%	50%	100%	0%
B	II	5	Writing	1	1	1	1		0	1	1		0	1	1	4,0	-	4,0	2,0	67%	33%	100%	0%
B	II	6	Writing	1	1	1	1		0		0	1	1	1	1	4,0	-	4,0	2,0	67%	33%	100%	0%
C	II	7	Writing	1	1	1	1		0	1	1		0	1	1	4,0	-	4,0	2,0	67%	33%	100%	0%
C	II	8	Writing	0	1		0	0	1		0		0	1	1	1,0	2,0	3,0	3,0	50%	50%	33%	67%
D	III	9	Writing		0		0		0		0		0		0	-	-	-	6,0	0%	100%	0%	0%
D	III	10	Writing	1	1	1	1	1	1		0	1		0		4,0	-	4,0	2,0	67%	33%	100%	0%
D	III	11	Writing	1	1	1	1	1	1		0		0		0	3,0	-	3,0	3,0	50%	50%	100%	0%
D	III	12	Writing	1	1	1	1	1	1		0		0		0	3,0	-	3,0	3,0	50%	50%	100%	0%
E	III	13	Writing	0	1	1	1		0		0		0	0	1	1,0	2,0	3,0	3,0	50%	50%	33%	67%
E	III	14	Writing		0		0		0		0		0		0	-	-	-	6,0	0%	100%	0%	0%
F	IV	15	Writing	1	1		0		0	1	1		0	1	1	3,0	-	3,0	3,0	50%	50%	100%	0%
F	IV	16	Writing	1	1	1	1		0	1	1		0	1	1	4,0	-	4,0	2,0	67%	33%	100%	0%
F	IV	17	Writing	1	1	1	1		0	1	1		0	1	1	4,0	-	4,0	2,0	67%	33%	100%	0%
			Multiple choice in context	13	15	12	12	4	5	7	7	2	2	8	9	46	4	50	52	49%	51%	92%	8%

APPENDIX VI

Participants' grammar tests.